Response to reviewers

A Machine Learning Examination of Hydroxyl Radical Differences Among Model Simulations for CCMI-1

Nicely et al.

On behalf of all coauthors, I thank the reviewers and commenters for their time taken to read the manuscript and offer constructive comments. They have served to significantly strengthen the analysis. Below, we address each comment and, where applicable, detail how the manuscript was revised in response. Original reviewer comments are shown in black font, and our responses are shown in blue.

Reviewer 1 – Dr. Peer Nowack

The paper by Nicely et al. uses a neural network approach to infer drivers of differences in OH/methane liftetimes among chemistry-climate models. In addition, the approach is used to understand modelled historic trends and variability in these variables. The method itself has been applied in similar form before (cf. Nicely 2017), but here it is applied to a novel set of specified dynamics CCMI simulations.

Overall this paper is a nice example of how machine learning can be used to provide novel insights into chemistry-climate model differences and I enjoyed reading it. I would therefore definitely recommend rapid publication subject some revisions and clarifications concerning my comments listed below. Major comments:

- The use of neural nets and especially their cross-validation requires further motivation and explanation. I know this can feel like unnecessary repetition to the authors given that the method has been described previously, but it is an essential aspect due to the central role of the method here. For example, when I first read the paper I was entirely unclear if all results might be subject to overfitting and if the sampling was done in space or time as well as how the data was split into training, cross-validation and test datasets; an essential aspect of any machine learning application. I now understand from reading the other paper that probably regressions were fit on an 80%/10%/10% split of the year 2000, using each grid cell as one sample for a month (rather than samples being ordered by time). Is this still valid? Is early-stopping really the only method you used to manage the bias-variance trade-off? This point is particularly important as evaluation results are given only for the year 2000, which as mentioned is used for training. Given that the year was used for training it would not be surprising if the neural net can fit the data almost arbitrarily well if overfitting wasn't sufficiently counteracted. Maybe show results/evaluate for all years that you did not use for training? I would also explicitly mention the sample size for each dataset (all models are interpolated to the same resolution?).

We agree with Dr. Nowack that the details of the method are important and have included more description in response. Specifically, we now write in the main text, at **L169**:

> "Each model gridbox located below the tropopause (thermal, following the WMO definition, for all models except GEOS Replay, which uses a "blended" tropopause calculation combining thermal and

> potential vorticity definitions) is a single sample, so sample sizes are determined by a model's vertical and spatial native resolution. The number of tropospheric model grid points, and thus the training dataset sample size, is indicated for each model in Table S1 and always exceeds 100,000. Because separate NNs are trained for each month, and monthly mean output from each model simulation is used as input and training data, the dataset does not represent diurnal variations in OH chemistry."

and at **L205**:

> "For training, the model output is randomly split 80%/10%/10% into training, validation, and test datasets. During that process, the data from the training set is used to actively adjust weighting factors, and the validation set is evaluated to determine a training stopping point. When errors in predicting the validation data grow after adjusting weighting factors some number of iterations in a row, it is determined that the NN model prior to the growth in errors likely reached a local minimum in its cost function. This manner of "early stopping" helps to prevent over-fitting, though application of the NNs to alternative years is not immune to over-fitting, an issue discussed further in Section 4.3.1. For further application of this method across varying time scales, we would recommend a more methodical approach to sampling model output in time as well as in space. The final 10% of data is then used to independently test the resulting NN, and compare between different training iterations. A total of five trainings were performed for each NN, and the NN with best performance (evaluated by the correlation coefficient from comparison of NN-calculated and model-simulated OH values) was chosen as the NN to be used in further analysis."

To further detail the training datasets used here, we have added Table S1 to the Supplement, listing the sample size of the datasets (i.e., number of tropospheric model grid points) used for each model, for each month.

We now include performance metrics of all NNs relative to year 2000, as were used in choosing between training versions. Figures S1-S4 have been added, showing correlations of NN-calculated and model-simulated OH, and new Tables S2 and S3 provide statistics of each NN used in this analysis. Associated text is included at **L300**.

We have also added evaluation of the NNs' performance for years other than 2000, and have modified our analysis of the time series of CH_4 lifetime based on this more quantitative identification of ill-performing (i.e., overfit) NNs (whereas before, we had subjectively removed a few cases of "spurious results" that stood out by eye). We use the evaluation of NN performance for each year as a guide; if the r^2 value of the NN-calculated OH (compared to the native model's OH) is greater than some threshold for a given year, then we will use that NN for that year. If not, that NN will be excluded, for that year. We somewhat arbitrarily decided on an r^2 threshold of 0.95, though we found the resulting variations and trends in CH_4 lifetime to be relatively insensitive to varying this threshold within reason. To demonstrate this, we have added the original Figs. 8 and 9, generated without the new quality check, to Supplement as Figs. S24 and S25.

The implementation of this NN quality check changed our results slightly (the trend in τ_{CH_4} due to tropospheric O_3, for instance, is a bit larger in magnitude), but the major conclusions of the analysis remain unchanged.

We thank the reviewers for suggesting more attention to NN overfitting, and have revised the manuscript at **L605** to describe our modified approach:

> "These examples of spurious results highlight an issue that must be treated with caution when using machine learning approaches. Because the application of our NN method to time series analysis is an extension beyond the originally intended purpose, not all NNs are sufficiently generalizable to reliably reproduce OH for years other than the training year, 2000. To account for this, we evaluate each NN for all years by inputting variables from each year. With this test, all inputs are changed, not just a single input at a time. The resulting OH, as depicted in Figures S16-S23 for select years, compares well to the native model's OH field for that year in many cases, but not in all. Considerable bias occurs at low OH mixing ratios, though we note that near-zero concentrations will likely not affect the resulting globally-integrated τ_{CH_4} unless values are grossly overestimated. This evaluation also represents a rigorous test of the NNs, as significant shifts in numerous inputs at once might push the NN algorithm into new phase space not encountered during training, much more so than only changing one input at a time, which is our approach in the subsequent time-series analysis. Nonetheless, we limit the influence of poorly generalizable, or "overfit," NNs by only including in the multi-model mean results for the years in which a NN reproduces its native model's OH field with an r^2 value greater than or equal to 0.95. For four NNs (one per month) created for each of 8 CCMI models, across 36 years, the potential application of the NNs to 1152 calculations (4×8×36) is reduced to 696 calculations using this test. Results from this point forward are subject to this quality check, and were found to be insensitive to the r^2 threshold imposed. This insensitivity is demonstrated by alternate versions of the figures to come, placed in Supplement, generated using all NNs rather than the quality-filtered NNs."

- I would like an additional explanation of why neural nets were used in the first place. I know they can model complex non-linear functions (which is one point that could be mentioned), but there are many algorithms that can do the same but would probably be more suited for inference tasks such as the one attempted here. Random forests, for example, would immediately provide feature importances for the regression models themselves and it would be easier to test dependencies between correlated variables (e.g. ozone, T, humidity) where it is unclear what is cause and effect. I do not ask for a refit with different algorithms, but it could be mentioned in terms of future work/context.

We started this work, as proof-of-concept, ~year 2013 and, since we then had the framework in place to conduct the analysis, we largely adhered to our original method. At that time, we were not aware of the random forest regressions approach, though we have since learned of the technique's benefits, including the feature importance capability. In the event that we are able to continue work in this area in the future, we view it as a high priority to explore the use of alternate techniques, though we remain confident that neural networks are suited to modeling the non-linear aspects of atmospheric chemistry when approached with appropriate caution and quality control measures.

To ensure the reader is aware of these alternative approaches and their similar suitability, we have added text at **L223** as follows:

> "We note that alternative machine learning algorithms have seen increased application to problems within atmospheric science in the last few years, and may be equally or even better suited than neural networks to studying non-linear chemical systems. In particular random forest regressions and gradient boosting techniques offer greater computational efficiency and, in the case of random forests,

> have the capability to quickly identify which inputs are most strongly influencing the calculated output, known as "feature importance" (Hu et al., 2017; Keller and Evans, 2019; Liu et al., 2018). …As such, we encourage exploration of …additional algorithms for future machine learning applications to atmospheric chemistry."

We have also added text further justifying our use of neural nets, as suggested, at **L103**:

> "NNs in particular are capable of modelling complex non-linear functions, making them a suitable technique for studying the non-linear chemistry involved in OH production and loss."

- some more reflection on the role of the nudged dynamics: the authors mention that one of the reasons why temperature is less important in explaining inter- model differences is the fixation to a common atmospheric background state by nudging. Alternatively, correlations with other variables such as ozone are offered as an explanation. Could the same not be said about water vapour? Maybe this would also explain why it is suddenly so much more important (relatively) to explain variability? What did you observe in this respect for the free-running simulations?

We have added text to the Model Simulations description, stating that most of the specified dynamics simulations do not constrain their water vapor to reanalysis data. While one would think these models would generally calculate water following temperature (which is constrained) as described by the Clausius-Clapeyron relation, this does not seem to be the case, as water appears as a "medium" driver of inter-model differences (4th in the ranking on Figure 5). It's likely that even small differences in water have a large effect on τ_{CH_4} due to its important role in primary production of OH.

In the manuscript, we had previously referred to the results of the free-running simulation to interpret this point, at **L418**:

> "We note that T differences between the SD simulations are likely limited due the meteorological constraints imposed on the models. However, examination of the free-running simulations, discussed in the Supplementary Material, also shows practically no impact of T on OH. Thus, we conclude that the effect of temperature on OH chemistry is likely indirect, acting through pathways embodied by other variables, such as H_2O and species that exhibit strongly temperature-dependent reaction rates."

We have also added a note regarding Dr. Nowack's point that H_2O is relatively more important in explaining temporal variability at **L644**:

> "It is interesting to note that H_2O plays a stronger role in the overall temporal trend of τ_{CH_4}, as compared to its role in explaining inter-model differences. This is likely due to the fact that temperatures were constrained in the specified dynamics simulations, which in turn should determine the water vapor calculated within the models."

- the randomness of neural networks: it seems that only one network is fit per model. Unfortunately, neural networks behave somewhat randomly, which is essentially the result of many different local minima in the cost function that can be found during the weight optimization process. Therefore, I would expect that the networks for each model would already be different due to different random initializations of the networks even if the

> chemistry models would be identical. I would strongly encourage the authors to test the relative importance of this randomness aspect compared to the actual inter-model differences. For example, they could train five-ten neural networks for two of the models (subject to an objective optimization/early stopping procedure) and show the spread in the results when these different network realizations of the two models are compared (instead of only one realization for each). No need to get started with different network architectures, which would similarly affect the results, I assume.

We trained five NNs initially for each model, and choose from among those the top performer, as we now explain in the text (see our response to the first comment, above). We have performed the suggested analysis, reproducing Figure 4 from the main text for five trainings of the GMI and OsloCTM NNs. These reproductions, using different NN versions, have been added to the Supplement as Figures S5 and S6. Visual, qualitative comparison of all 5 versions reveals very similar spatial distributions and magnitudes for the most part, though the values of the change in CH_4 lifetime do vary modestly, depending on the variable. For instance, the standard deviations in $\Delta\tau_{CH_4}$ from JO^1D and HCHO swaps into the GMI NNs are about 0.2 years, though some of the NNs included were not as highly-performing as the chosen NN.

In any case, we agree with Dr. Nowack on the importance of the reader being aware of this fact, and have added the following text in our discussion of Figure 4, at **L363**:

> "A fourth issue is the fact that NNs can exhibit some degree of random behaviour, based on how they were trained and initialized. Our method involved training 5 NNs and selecting from those the one that performed best when compared to the independent test dataset. That single NN was used in all subsequent analysis. However, it is a useful exercise to evaluate the role of NN randomness in our results. We show in Figures S5 and S6 the left and right panels of Fig. 4, reproduced for the alternate NN trainings of the GMI and OsloCTM models, respectively. A visual comparison of tropospheric OH column differences among the five trainings of each model's NN reveals markedly similar spatial distributions and magnitudes. The values of calculated $\boldsymbol{\Delta\tau_{CH_4}}$ do differ somewhat between the training instance, with larger effects on some variable swaps than for others. For instance, the standard deviation of the values of $\boldsymbol{\Delta\tau_{CH_4}}$ calculated for all five trainings of the GMI NN is about 0.2 years for the $J(O^1D)$ and HCHO swaps, but less than 0.05 years for O_3 and NO_x. We note, though, that some of the NNs displayed in Figures S5 and S6 exhibit worse performance than the one ultimately chosen for subsequent use. As a result of this exercise, the uncertainties resulting from this analysis method may be considered, at most, to be ~0.2 years."

Minor comments:

- p. 4, l.107-109: revise second part of the sentence.

We have changed this line from "…seeks to further inter-model evaluation…" to "seeks to enable inter-model evaluation…"

- p. 4/5; model simulations: since UV fluxes and stratospheric ozone are discussed maybe briefly mention if all/which models include interactive stratospheric chemistry, or how it is treated otherwise.

We have included the statement at **L137** that "All models, here and including those described below, include interactive stratospheric chemistry."

- section 3.1 I think there should be more detail here; essentially another small subsection on the cross-validation method.

Please refer to our additions to the text described under the first major comment, above. We believe this adds considerably to the detail concerning our method, as Dr. Nowack requests.

- p. 5 l. 161: 'mutually exclusive' - what do you mean by that here?

We have changed the language here to "outside the ranges," to improve clarity.

- l. 165-170: Maybe try a variation of the input features? The cross-correlations are indeed an obvious problem for the interpretation. Did you consider fitting two different networks, e.g. one with JO1D, one with column ozone and consider how well they do on the cross-validation dataset? I am also wondering how these different networks would perform in different atmospheric regimes, e.g. column ozone being more important in the upper troposphere. JO1D (including clouds) becoming relatively more important in the lower troposphere? Can a single network for all grid cells capture these different regimes appropriately?

We did evaluate inclusion of just JO1D (and not column ozone, which was the method of our original "proof-of-concept" study in Nicely et al., 2017), use of just column ozone (and not JO1D), and inclusion of both at the time of original development of the method. While the first and third options were quite comparable in terms of NN performance, the second was not as effective, likely because overhead ozone is so far removed from the *in situ* OH quantity. JO1D, on the other hand, is the immediately-relevant measure of the UV light affecting OH at a particular time and place. Over the years we have worked on this, there has been a substantial amount of experimentation with the input variables chosen, and the set of inputs were determined to provide the best balance between strong performance on independent test data following training, and reasonable results following the inter-model swapping of variables (use of absolute values of CH_4 would result in non-sensical output, for instance).

To encourage the reader to consider the importance of architecture/input choice testing, in the case that they attempt a similar analysis, we have added text at **L229** stating:

> "We also do not intend to suggest that our chosen NN input list, architecture, and general method is the best approach; input variables were largely determined by available output, and architecture testing was conducted on the computing resources available at the time of the study."

- l.228: performance for the year 2000 is strong – but this is the training year. Should the goal not be to evaluate on out-of-sample years. Maybe show an error plot for all years? I assume it gets worse the further one moves away from the training year, partly due to the extrapolation error?

We refer to our response to the first major comment, where we added evaluation of all NNs for years other than 2000 and adapted our time series analysis to remove poorly-performing NN instances.

- general remark on the extrapolation issues: could you give an estimate of how often you had to correct values in this way for each comparison/model (e.g. percentage o cases depending on the year)? This would give the reader a better impression of how important this factor is when considering the results. In addition, did you ever test how linear/non-linear the regression relationships really are? Maybe linear regression algorithms such as Lasso/Ridge would actually circumvent all these issues by being able to extrapolate better and still extract feature importances in a sophisticated enough manner (the resulting regressions would also be easier to interpret).

The extrapolation control method we developed is only utilized in the inter-model comparison portion of this analysis. Our reasoning for doing so is that any given model should simulate generally comparable conditions from year to year, aside from regime shifts in anthropogenically-emitted species and strong ENSO events that shift locations of convection, biomass burning, etc. And, if our NNs are moderately generalizable, then small excursions outside the range of variable values on which the NNs were trained should be manageable (which we found was not always the case and have now accounted for, as explained under the first comment, above).

We still agree with Dr. Nowack, though, that some indication of how frequently extrapolation control is employed in the inter-model comparison analysis would be informative. To do this, we have written out flags for instances in which adjustments to swapped input variables are made. We separate these instances into “coarse” and “fine” adjustments, the former describing the case when an incoming value falls completely outside the range of tropospheric values from the NN’s native model, incurring a presumably large adjustment, and the latter describing smaller changes made to conform the other model’s variable to the native model’s chemical regimes. For the January NNs only (the extrapolation control code is rather inefficient; it took several weeks to generate these results alone), the percentage of cases (total number of tropospheric grid points) that undergo coarse adjustment are 3.5% on average, while cases in which fine adjustments are performed average 18.8%. A large number of the fine adjustment cases are driven by inconsistencies in CH_4, though individual models may have other factors that contribute significantly.

The result of performing these adjustments is to dampen the calculated impact to OH and τ_{CH_4}. However, the analysis already reveals instances with large changes to τ_{CH_4}, so we think it is appropriate to state more conservative results with a higher level of confidence.

We have added text to inform the reader how often extrapolation control is invoked at **L248**:

> “For reference, we tally the number of instances in which extrapolation control is invoked for two categories: coarse adjustments, when a NN input value from another model falls entirely outside the range of the NN input values from the native model, and fine adjustment, when a value from another model must be tweaked to preserve the native model’s chemical regimes. On average, coarse adjustments are incurred for 3.5% of all swapped data points, while fine adjustments are made to

> 18.8% of swapped values. We find that extrapolation control is critical to achieve meaningful results with the NN inter-model comparison method, though it necessarily forces the attributed changes in OH and τ_{CH_4} to be conservative estimates."

We have not explored the use of linear regression techniques such as Lasso/Ridge in order to ameliorate issues pertaining to extrapolation, but we would like to encourage the reader to do so. We have added text at **L228**:

> "Additionally, linear regression algorithms such as Ridge and Lasso regression may be beneficial in curbing issues related to extrapolation."

- l. 484: maybe I approach this one too naively, but why would I expect to model a CH4 trend if CH4 is normalized by its maximum value in each year? I assume the maximum value shows a trend somewhat proportional to the average trend?

Dr. Nowack is correct on this point; we cannot attribute any meaning regarding the "CH_4 feedback factor" to the trend in τ_{CH_4} due to CH_4 found here, as a result of CH_4 being normalized. We have taken steps to remove language suggesting that a trend in OH due to CH_4 is found, emphasized in some figures that CH_4 is normalized by using the notation "CH_4^{NORM}", and removed the trend due to CH_4 data point from our final figure, comparing the CCMI model trends in τ_{CH_4} to our previous empirical study's trends in $[OH]^{TROP}$, since they are not comparable.

We chose to leave "CH_4^{NORM}" in Figs. 7-9 because there is *some* meaning in this value; it represents changes in the spatial and vertical distribution of CH_4 within the troposphere, which, based on how the CH_4 collocates with high OH concentrations, can influence the resulting τ_{CH_4} value. We have added explanation to this effect at **L497**, where we state:

> "Because we are relying on the same NNs used for the inter-model analysis, we emphasize that the CH_4 fields used here are still normalized, separately for each year. As a result, the variations in τ_{CH_4} due to CH_4 should not be interpreted as a measure of the CH_4 feedback factor (Prather et al., 2001). Instead of representing the change in OH with a change in absolute concentration of CH_4, the numbers shown here signify the change in OH with a change in how CH_4 is distributed within the atmosphere. Largely, one would expect this to remain constant over time, though results from this analysis of the CCMI simulations suggests there are some modest changes in τ_{CH_4} attributed to the distribution of tropospheric CH_4. Should a similar method be applied to analysis of temporal variations in OH in the future, we would encourage training the machine learning algorithm on data spanning all years such that use of CH_4 absolute values would be possible."

Reviewer 2

Nicely et al. (2019) attributed OH differences among CCMI models into a number of parameters using a neural network approach. They found the major drivers for the decline in methane lifetime are tropospheric O3, JO1D, NOx, and H2O, with CO contributing to the OH interannual variability. It is a very interesting study with very popular machine learning technique. The manuscript is in general well written and well organized. I recommend acceptance of the manuscript after addressing below questions.

Neural network setup

As described in the manuscript, one NN is trained for each model for each month and all the training is performed for year 2000. So how is it applicable for the input with a lengthy period? Some variables would undergo significant changes from the 1980s to 2010s. What if the NN trained for year 2000 is not suitable for 1980s or 2010s?

This is certainly a concern, and we have added new analysis that identifies how well the NNs, trained on output from year 2000, perform for other years. The majority of NNs continue to perform strongly, though a number of them encounter conditions that cause large deviations in their predicted OH.

We refer the reviewer to our response to Dr. Nowack's first major comment, above, for an in-depth description about how the manuscript has been changed to address this point, including adjusted analysis, new methodological details in the main text, and many new figures in the Supplement.

There is one concern that when you substitute a single input taken from one model into another. Would this affect the original chemical regime or atmospheric condition? Would there be some "relaxation" in the system to approach original condition? In that sense, it could reduce the sensitivity of OH to the differences in the input.

We agree with the reviewer, that this analysis neglects "true atmospheric behavior," as you might call it – feedbacks and relaxation effects are ignored as we instead are calculating an instantaneous change that would occur if you could magically perturb a single species or variable. But we would still regard this as a useful exercise to both parse the main influencers of OH chemistry and identify the causes of inter-model differences, which are difficult to do otherwise. We have added text acknowledging this point at **L476**:

> "A final qualification is this analysis constitutes a foundationally hypothetical experiment. It essentially addresses the questions, "What if we could instantaneously switch the fields of just one chemical species between two global models? What would be the impact on OH? on $\boldsymbol{\tau_{CH_4}}$?" This approach, then, necessarily neglects the roles of feedbacks in the atmospheric system (e.g., if the NO_x field is perturbed, this will propagate to changes in O_3 as well, with time). However, for the objective of teasing apart the influences on global OH abundance and $\boldsymbol{\tau_{CH_4}}$ and explaining inter-model differences, a notoriously difficult task, we regard our approach as a valuable exercise.“

Lastly, it is more of a broad question. To what degree that the trained NN can realistically represent the non-linear chemical system. In this work, there are a number of variables are input to the NN. The weighting factors can be adjusted during the training process, but if there are more inputs or different inputs, the weighting factors could be different? Would this affect conclusion? How to deal with this issue?

Neural networks are generally highly capable of modeling non-linear functions, which is the main reason why we chose this approach originally. Within the NN architecture, we use hyperbolic tangent activation functions, which are non-linear. As these functions are used many times over, in parallel, they enable complex fitting of multi-dimensional, non-linear surfaces. The reviewer is correct, though, that once the training process is complete, the weights of our chosen NN are fixed, and the insertion of different inputs can cause issues. To deal with this, we implement the "extrapolation control" method described in the text. In our revisions, we have also taken the extra step of evaluating our NNs across all years, to exclude NNs that do not generalize well (i.e., reproduce well the OH for a particular year, presumably due to some new conditions encountered) from our analysis.

Specific comments:

Page 4, 121-125, is water vapor nudged for all the REF-CISD simulations? If not, what are the REF-C1SD simulations that nudge water vapor?

We have added statements to the text describing which models include specific humidity nudging in their specified dynamics schemes, at **L131**:

> "Particularly relevant to this analysis is the nudging of specific humidity, which is only performed in the MOCAGE model, of the models we analysed."

and **L156**:

> "All CTMs except GEOS-Chem calculate water vapor interactively in the troposphere. GEOS-Chem instead uses specific humidity fields from the MERRA reanalysis."

Page 8, line 244, but also over tropical ocean?

Yes, it is true that changes to O_3 and NO_x influence OH over the tropical oceans as well as over the continents, though the maximum changes in OH appear over land. We have changed this statement to read, at **L319**:

> "…exert the greatest influence on OH over the climatological tropics, with maximum impacts over land but extending over the oceans as well."

Page 9, line 261-264, could you elaborate "buffering effects"?

The buffering effects we are referring to mostly involve more complex hydrocarbon chemistry, and so we have changed this example from CH_4 to isoprene. The text now reads, at **L337**:

> "For example, one model may be sensitive to an increase in isoprene, causing OH concentrations to drop in response. Another model may incorporate buffering effects (such as reactions involving oxidized volatile organic compounds (Lelieveld et al., 2016; Taraborrelli et al., 2012) that allow OH to be recycled…"

Page 9, line 278-282, this is similar to "relaxation" that mentioned in the general comments.

The example given at this location, regarding the implementation of extrapolation control preventing a large change in CH_4 from being conveyed to the NN, does not exactly represent a "relaxation" of the system, but rather, a logistical issue with the method, preventing our even testing a large perturbation in the NN out of an abundance of caution. In general, though, we agree with the reviewer that the issue of relaxation is not directly addressed by our method, and so the text added at **L476** (described in our response to the general comment, above) discusses this point.

Page 11, line 326 &ff, Figure 5, the impacts of temperatures are small due to the specified dynamics in the model. What about water vapor? If specified water vapor is also imposed, are the impacts of water vapor still large? You may want to check the models with the specified water vapor.

This is a very interesting idea, though the only two models that used specified water vapor (MOCAGE and GEOS-Chem) used two different reanalysis data sources (ERA-I and MERRA, respectively). The mean $\Delta\tau_{CH4}$ values due to H_2O for these two models are quite different, though there is at least overlap in the 1σ about the mean, represented by the "whiskers" in Fig. 5. Had there been more models imposing a water vapor constraint, from the same reanalysis data, this would be worth exploring further.

Page 11, line 336-378, what do you mean by "reminder term"?

We describe the "remainder term" three paragraphs prior to the location identified by the reviewer, where it is instructive to refer to Table 1. To improve clarity, we now include the text: "…the remainder term (or term labelled "Mech." in Table 1)…" at **L428**.

Page 16, line 487-489, are you talking about latitudinal gradient or vertical distribution?

Because of use of normalized CH_4 prevents inferences regarding the true "CH_4 feedback" on OH, we have removed the discussion at Pg. 16, line 487. However, new text placed earlier in the discussion (starting **L497**) also refers to the "distribution of methane." We have included "both vertically and spatially" as clarification, as changes in the collocation of OH and CH_4, no matter where in space, could impact the calculated τ_{CH4}.

Reviewer 3 – Dr. Leif Denby

I am only commenting on the machine learning aspect of the submitted manuscript. Apologies for overlooking for not providing more general feedback.

1. In section 3.2 I would rephrase the sentence containing "mimic the tropospheric chemistry" to include "predict the instantaneous OH" concentration. As is written now it might give the impression that the time evolution is predicted by the neural networks as the research presented is about reactions.

We agree with Dr. Denby on this use of language and have made the recommended change.

2. I find the sentences "Briefly, one NN is trained for one model, for one simulation month at a time." and "To reduce computational demands, we establish NNs for four months, one for each season..." a little contradictory. Is training done on one month or on four months of input? How is it possible to do both? It might be that the reader should simply study the referenced paper, but I find this a little unclear.

We apologize for the confusing language here; training is done on one month of input; we generate separate NNs for each month that we look at; and we look at 4 months. We have attempted to clarify this by changing the text to read, at **L166**:

> "Briefly, four NNs are trained for one model, each for one simulation month. To reduce the computational demands of NN training, we only establish NNs for four months, one for each season…"

3. It would be nice to a brief comment on why models were trained for each month separately. Was this done because the temporal variability couldn't be captured by a single model? Does the skill of each model vary through the month? I assume that at the ends of the month (where there is transition between which model is used) there might be a reduction in skill. But maybe the predictions match seamlessly when switching between models.

In our early years of developing this method, we encountered a couple of issues that resulted in our decision to only train an NN for a single month. When we first attempted to ingest all model output for an entire year into the Matlab NN software, on which we still rely, and on a graduate student's laptop (albeit a powerful one – on which we do *not* still rely), we unsurprisingly encountered memory issues when attempting to train the NN. Then, during our limited attempts to randomly sample model output across all months to generate a training dataset, we found that the NNs did not perform well.

Now that considerably more progress and application of machine learning to scientific questions has taken place, we would encourage a more methodical and strategic sampling of the model domain to create a training dataset. To ensure the reader is aware of this point, we have added text at **L231**:

> "It is possible that a single NN could suffice for predicting OH variations throughout an entire year, rather than for just a single month, following methodical subsampling methods to create the initial training dataset."

Regarding questions about how the NN performs "through the month," we use only monthly mean output from the CCMI models examined here (monthly mean output is commonly what is made available from these large model intercomparison projects). So, we are not able to address issues of varying performance throughout a simulated month.

4. The "Inter-model comparison" is nice. With the restriction on the numerical range of the values which are substituted I feel that feature importance could similarly be inferred by simply shuffling (across time) all values for a specific feature, similarly to how it's done for random forests. Is there a reason why this wasn't attempted here? Isn't there a concern that using the presented method that one might infer low feature importance for fields that simply vary little between models?

We thank Dr. Denby for the statement of support. His follow-up questions relate to the ultimate goal of the analysis. In our case, the objective is to explain why global models of atmospheric chemistry give different quantities of global mean OH and τ_{CH_4}. In that case, we are less concerned with quantities that are quite consistent among the models, even if they have the capability of strongly altering OH chemistry. We go into this a bit in the Discussion (~**L354**), explaining that two conditions must be met to incur a change in OH: differences in the input between the two models, and sensitivity of OH to that input.

The questions posed above would be interesting to address if one were examining which inputs have strong "feature importance," which is not necessarily the goal of the inter-model comparisons, but is, in essence, what we have done in the time series evaluation.

Short Comment 1 – Mr. Karl M. Seltzer, Dr. Prasad Kasibhatla

General Comments

The manuscript "A Machine Learning Examination of Hydroxyl Radical Differences Among Model Simulations for CCMI-1" by Nicely et al. discusses a topic that is of high interest to the Atmospheric Chemistry and Physics community. Possibly the most perplexing issue in atmospheric chemistry is the unexpected stabilization of global methane concentrations from ~2000-2006. This study attempts to unravel the individual CTM drivers of the hydroxyl radical in a suite of simulations, thus illuminating the changes, and reason for said changes, in the primary termination pathway for methane, as simulated by each CTM.

While this work is important, we do have concerns about how some of the results are presented and methods are employed in this analysis, both of which constitute major comments. We will describe both in more detail below, followed by some minor comments.

Major Comments

1. In Figures 7-10, results from the CH_4 signal, as it relates to changes in tropospheric OH, are presented. While the text does explicitly state that "CH_4" is a normalized value based on the maximum tropospheric value, we believe the presentation of the results in Figures 7-10 and much of the language used throughout the manuscript can lead to substantial confusion on the part of the reader. The reader might reasonably interpret the results as an estimate of the sensitivity of τ_{CH4xOH} to changes in CH_4 abundance (i.e. the CH_4 feedback factor). One example: the inclusion of CH_4 in Figure 10 makes a comparison of the "CH_4" value reported in this study (i.e. NOT the CH_4 feedback factor) with the calculated CH_4 feedback factor from Nicely et al., 2018.

   Based on our interpretation of the methods employed here, the authors did not analyze the CH_4 feedback factor. Since it seems the better characterization is that the global *distributions* of CH_4 concentrations were analyzed, we think the authors need to re-write any discussions related to CH_4 results throughout the manuscript to make this distinction abundantly more

> clear, and should possibly remove the characterization of "CH_4" in Figures 7-10. Similarly, it is not clear why CH_4 concentrations were normalized. Presumably, the same analysis using non-normalized values of CH_4 would be able to capture the CH_4 feedback?

We fully acknowledge that the impact of CH_4 on the trend in $\boldsymbol{\tau_{CH_4}}$, as it is found here, does not represent the CH_4 feedback factor. This was a late realization, and some of the language and figures in the manuscript may have been misleading as a result. We have taken steps to remove this misleading content in the following ways:

- During early discussion of the "Time series evaluation" results, we attempt to present this issue in a forthright manner. Starting at **L497**, the text now reads:

  > "Because we are relying on the same NNs used for the inter-model analysis, we emphasize that the CH_4 fields used here are still normalized, separately for each year. As a result, the variations in $\boldsymbol{\tau_{CH_4}}$ due to CH_4 should not be interpreted as a measure of the CH_4 feedback factor (Prather et al., 2001). Instead of representing the change in OH with a change in absolute concentration of CH_4, the numbers shown here signify the change in OH with a change in how CH_4 is distributed within the atmosphere, both vertically and spatially. Largely, one would expect this to remain constant over time, though results from this analysis of the CCMI simulations suggests there are some modest changes in $\boldsymbol{\tau_{CH_4}}$ attributed to the distribution of tropospheric CH_4. Should a similar method be applied to analysis of temporal variations in OH in the future, we would encourage training the machine learning algorithm on data spanning all years such that use of CH_4 absolute values would be possible."

- In Figures 7-9, we now label the time series/trends due to CH_4 as "CH_4^{NORM}" to serve as a reminder that the CH_4 with which we performed the analysis is normalized. We chose to leave "CH_4^{NORM}" in Figs. 7-9 because there is *some* meaning in this value; it represents changes in the spatial and vertical distribution of CH_4 within the troposphere, which, based on how the CH_4 collocates with high OH concentrations, can influence the resulting $\boldsymbol{\tau_{CH_4}}$ value.
- We have removed entirely the CH_4 data point in Figure 10 comparing the CCMI model trend, as evaluated by NN, to the Nicely et al. (2018) trend, and all discussion associated with it, as the two values do not provide an "apples to apples" comparison.

Regarding why normalized CH_4 was used in the first place, we sought to utilize the same NNs trained for the inter-model comparison application for the new analysis of OH time series. Absolute values of CH_4 mixing ratio as NN inputs were initially attempted for the inter-model comparison, but yielded non-sensical results since the models calculated very different CH_4 fields in some cases. In the case of CCMI, the models are fairly similar, since they use the same CH_4 boundary condition, but the external models that did not formally participate in CCMI still pose the same problem.

In the case that we could dedicate considerably more time to this work, we would ideally train new NNs for the time series analysis portion of the project using absolute CH_4 mixing ratios. This would necessitate that we create a training data set consisting of samples across all years, lest we run into the same dilemma of having the NN trained on a relatively narrow range of CH_4 values. This type of subsampling should be performed strategically, and, along with the actual training of the NNs, would be computationally demanding and require a substantial amount of time, thus we regard this as beyond the scope of our current manuscript.

2. The sensitivity of τ_{CH4xOH} to changes in CH_4 abundance reported by CTM studies are reasonably consistent and range from -0.25 to -0.35 (Prather et al., 2001; Fiore et al., 2009; Holmes et al., 2013, Holmes 2018). That is, the tropospheric OH abundance declines by 0.25%-0.35% for every 1% increase in CH_4 abundance (Prather et al., 2001). The IPCC AR5 reported that global CH_4 abundance grew by ~13% from 1980 to 2010 (Ciais et al., 2013).

   Assuming the models used here respond in a similar manner to other published CTM studies, the CH_4 feedback should have yielded a ~3.3%-4.6% decrease in tropospheric OH between 1980-2010 (or equivalently, 1.1%-1.5% per decade). That driver should theoretically be captured in the net results presented in Figure 6.

   As noted on Line 457, the mean downward trend in τ_{CH4} of Figure 6 is 1.8% per decade. Therefore, the residual (i.e. all of the other factors outside of the CH_4 feedback) should be ~(-1.8% - 1.3%) → -3.2% per decade (note: 1.3% is the average of 1.1% and 1.5%). This is much larger than the ~residual of -1.9% reported on Line 457 (~residual because it does not include the CH_4 feedback factor). Therefore, since the τ_{CH4} budget does not appear to be closed when adding up all of the variables (including the CH_4 feedback), this suggests that the methods used here have difficulty in deriving the contributions of individual drivers. If so, that would be a fundamental issue with the methods used to derive Figures 7-10. Here are some ways we believe the authors can build confidence in the methods used here:

   a. A quick first step would be to add up all of the components for each model in Figure 7 and plot their change, side-by-side, to the values presented in Figure 6 (normalized to 2000 values for consistency). Do the trends match? If yes, since the NN method does not account for the CH_4 feedback and CTMs are known to have a robust and consistent CH_4 feedback, why do they nonetheless match? If no, can the missing CH_4 feedback explain the difference?
   b. A lengthier, but maybe necessary test: Experiment with one of the CTMs. For example, re-run GMI with the year 2000 repeating for all variables, except CO. This might only be necessary for a few select years, such as 1985 and 1998. Do these results match the dark blue line in Figure 7e? One or two examples of these types of validation steps would really increase our confidence in the driver analysis.
   c. When attributing specific, individual drivers to trends, Random Forests are considered better machine learning tools (Grange et al., 2018). It is likely easy to swap out the NN code in your analysis with a random forest. Experiment with one of the models. For example, run the random forest algorithm for GMI's 2000 results and repeat the process for Figure 7. How different are the results?

We acknowledge that these suggestions by Mr. Seltzer and Dr. Kasibhatla would make a rigorous test for the application of our method to time series and determination of trends in OH/τ_{CH4}. We have taken steps to build confidence that the method is fundamentally sound following their item (a.) above.

Below we have created a table listing the overall trends in τ_{CH4}, taken directly from the CCMI models (i.e., Figure 6), the overall trend calculated by totaling each component from the NN analysis (Figure 7, excluding the obviously spurious cases discussed in the text: EMAC CH_4 and MOCAGE O_3 Column), and the implied CH_4 feedback factor found by subtracting the latter from the former.

Model	Native model trend in τ_{CH4} (Fig. 6) (% decade^{-1})	Summed trend in τ_{CH4} from NN-calculated components (Fig. 7)	Implied CH_4 feedback (Column 2 – Column 3)
CAM4Chem	–2.69	–3.15	+0.46
EMAC-L47MA	–1.22	–1.32	+0.10
EMAC-L90MA	–1.48	–1.50	+0.02
GEOSCCM	–0.70	–1.57	+0.87
GMI	–0.54	–1.86	+1.32
MOCAGE	–2.97	+1.32	–4.29
MRI-ESM1r1	–2.31	–2.34	+0.03
WACCM	–2.72	–2.97	+0.25

The values of the implied CH_4 feedback are all of the correct sign, except for MOCAGE, which generally demonstrates quite different behavior from the other models. The value for GMI is in good agreement with the 1.1-1.5 % decade^{-1} range that is cited in the comment above. We would identify a couple of issues with validating the method in this manner, though.

First, as described in our previous responses to reviewers, we now identify specific instances in which the NNs (for specific months and years) do not perform sufficiently well, and so the multi-model mean trend results that we show now use NN calculations "filtered" for only the high performing NNs. The numbers we quote in the above table, in Column 3, include all NN results except for the MOCAGE O_3 column and EMAC CH_4 contributions, and so admittedly include some dubious contributions. Because NNs for individual months are filtered out in the new quality-check, it would not be straightforward to calculate new $\Delta\tau_{CH4}$ values, on a year-by-year basis, for a single model. The aggregation of all models, into the multi-model mean results we present, allows us to assess the trends using results from all years.

Second, we do only perform this analysis for four months out of the year, so a truly apples to apples comparison with, e.g., CH_4 feedback factors from other studies would more aptly include all 12 months. And finally, as we now acknowledge following our response to Reviewer 2, there are secondary effects and "relaxation" that occur in the real atmosphere and in the global models we are examining, which the NNs may not capture. This analysis can instead be interpreted as evaluation of the instantaneous change in OH resulting from a hypothetical perturbation to a single chemical/radiative/physical variable. One should be aware that these perturbations often do not occur in isolation, though, and so we now caution the reader that the responses shown by this analysis may not be directly applicable to the real world.

Because we have chosen to remove discussion of our results regarding the trends in τ_{CH4} due to CH_4, we consider further analysis regarding the CH_4 feedback factor as beyond the scope of the current work. We do encourage further study of the issue, though, both by suggesting refinements to our method (i.e., creating a training dataset sampled across many years, to enable use of absolute CH_4 values in the NNs ~**L223**) and by endorsing movement away from CH_4 boundary conditions (which we believe hampers useful studies of the CH_4 budget with our present-day atmospheric chemistry models), toward interactive fluxes (**L687**).

Minor Comments

- Figure 3 compares the tropospheric OH columns from WACCM and the ANN-WACCM predicted tropospheric OH columns. As noted on Line 174, the training methods in this analysis were the same as those carried out in Nicely et al., 2017, which stated that the training/validation/testing datasets comprised 80/10/10% of all data. Therefore, it seems that 80% of the data that was used to construct the middle panel of Figure 3 was data that the ANN has seen before (i.e. from the NN training). Shouldn't this part of the evaluation be restricted to only the testing dataset?

While the actual evaluation of the NN post-training was performed on the testing dataset, we also wanted in Figure 3 to convey the spatial distribution and magnitudes of tropospheric OH column amounts, relevant for the interpretation of Figure 4. Since our NNs calculate OH on a 3-D basis, and we then integrate the columns in post-processing, generation of a similar figure showing only the 10% of model output used for training would not be possible (i.e., you likely wouldn't have a full vertical profile of OH for any single lat/lon coordinate).

We have, however, added considerable content displaying the performance of our NNs in the form of 2-D histograms in the Supplement, including for years other than 2000 (Figs. S16-S23). While we have newly adapted our analysis to rely only on the NNs that continue to show strong performance in reproducing its native model's OH for a given year, the overall conclusions have changed little as a result (see our response to the first comment by Dr. Nowack for further detail).

- In the paragraphs spanning Lines 423-448, there is a discussion about "spurious results". Are these results "spurious" just because they look out of place in Fig. 7, or are there some other quantifiable ways that might justify the label "spurious"?

We initially identified these "spurious results" by eye, but since looking more closely at the performance of our NNs across all years, we have instituted a quantitative threshold to determine when results from a particular NN/year should be disregarded. We choose a somewhat arbitrary r^2 threshold for this purpose, but we did test the effects of altering this threshold and found little change in our results (again, see our response to Dr. Nowack for further detail).

- Figure 9b: Don't CTMs have difficulty in capturing observation-derived estimates of IAV (Holmes et al., 2013)? That should be noted.

We have added text noting this in our discussion of Figure 9b, at **L637**:

> "The interannual variability of τ_{CH_4} is also calculated as the standard deviation of the detrended time series, shown in Fig. 9b, though it is relevant to note that CTMs have historically not captured the full interannual variability exhibited by observed OH proxies (Holmes et al., 2013)."

- Lines 482-498 should likely be removed. The comparison of the CH_4 results here and the CH_4 results in Nicely et al., 2018 are not an 'apples-to-apples' comparison, as noted by the authors in the sentence starting with "On one hand..." from Line 485.

We concur; this text has been removed.

References

Ciais, P., C. Sabine, G. Bala, et al., 2013: Carbon and Other Biogeochemical Cycles. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. 2013.

Fiore, A. M., Dentener, F. J., Wild, O., et al.,: Multimodel estimates of intercontinental source- receptor relationships for ozone pollution, J. Geophys. Res., 114(D04301), doi:10.1029/2008JD010816, 2009.

Grange, S. K., Carslaw, D. C., Lewis, A. C., et al.,: Random forest meteorological normalisation models for Swiss PM10 trend analysis, Atmos. Chem. Phys., 18(9), 6223–6239, doi:10.5194/acp-18-6223-2018, 2018.

Holmes, C. D.: Methane Feedback on Atmospheric Chemistry: Methods, Models, and Mechanisms, J. Adv. Model. Earth Syst., 10(4), 1087–1099, doi:10.1002/2017MS001196, 2018.

Holmes, C. D., Prather, M. J., Søvde, O. A. and Myhre, G.: Future methane, hydroxyl, and their uncertainties: Key climate and emission parameters for future predictions, Atmos. Chem. Phys., 13(1), 285–302, doi:10.5194/acp-13-285-2013, 2013.

Nicely, J. M., Salawitch, R. J., Canty, T., et al.,: Quantifying the causes of differences in tropospheric OH within global models, J. Geophys. Res., 122(3), 1983–2007, doi:10.1002/2016JD026239, 2017.

Nicely, J. M., Canty, T. P., Manyin, M., et al.,: Changes in Global Tropospheric OH Expected as a Result of Climate Change Over the Last Several Decades, J. Geophys. Res. Atmos., 123(18), 10,774-10,795, doi:10.1029/2018JD028388, 2018.

Prather, M., Ehhalt, D., Dentener, F., et al.,: Atmospheric Chemistry and Greenhouse Gases, in Climate Change 2001: The Scientific Basis. Third Assessment Report of the Intergovernmental Panel on Climate Change, 2001.

We thank Mr. Seltzer and Dr. Kasibhatla for these recommended references; we have added all but Ciais et al. (since our analysis does not quantify a CH_4 feedback and thus discussion of the CH_4 increase is less relevant) to our manuscript, where applicable.

A Machine Learning Examination of Hydroxyl Radical Differences Among Model Simulations for CCMI-1

Julie M. Nicely[1,2], Bryan N. Duncan[2], Thomas F. Hanisco[2], Glenn M. Wolfe[2,3] Ross J. Salawitch[1,4,5], Makoto Deushi[6], Amund S. Haslerud[7], Patrick Jöckel[8], Béatrice Josse[9], Douglas E. Kinnison[10], Andrew Klekociuk[11,12], Michael E. Manyin[2,13], Virginie Marécal[9], Olaf Morgenstern[14], Lee T. Murray[15], Gunnar Myhre[7], Luke D. Oman[2], Giovanni Pitari[16], Andrea Pozzer[17], Ilaria Quaglia[16], Laura E. Revell[18], Eugene Rozanov[19,20], Andrea Stenke[19], Kane Stone[21,22], Susan Strahan[2,23], Simone Tilmes[10], Holger Tost[24], Daniel M. Westervelt[25,26], and Guang Zeng[14]

[1]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA.
[2]NASA Goddard Space Flight Center, Greenbelt, MD, USA.
[3]Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore, MD, USA.
[4]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, USA.
[5]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD, USA.
[6]Meteorological Research Institute (MRI), Tsukuba, Japan.
[7]Center for International Climate and Environmental Research-Oslo (CICERO), Oslo, Norway.
[8]Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, Germany.
[9]CNRM UMR 3589, Météo-France/CNRS, Toulouse, France.
[10]National Center for Atmospheric Research, Boulder, CO, USA.
[11]Antarctica and the Global System Program, Australian Antarctic Division, Kingston, Australia.
[12]Antarctic Climate and Ecosystems Cooperative Research Centre, Hobart, Australia.
[13]Science Systems and Applications, Inc., Lanham, MD, USA.
[14]National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand.
[15]Department of Earth and Environmental Sciences, University of Rochester, Rochester, NY, USA.
[16]Department of Physical and Chemical Sciences, Universitá dell'Aquila, L'Aquila, Italy.
[17]Max-Planck-Institute for Chemistry, Air Chemistry Department, Mainz, Germany.
[18]School of Physical and Chemical Sciences, University of Canterbury, Christchurch, New Zealand.
[19]Institute for Atmospheric and Climate Science, ETH Zürich (ETHZ), Zürich, Switzerland.
[20]Physikalisch-Meteorologisches Observatorium Davos – World Radiation Center (PMOD/WRC), Davos, Switzerland.
[21]School of Earth Sciences, University of Melbourne, Melbourne, Australia.
[22]Massachusetts Institute of Technology, Cambridge, MA, USA.
[23]Universities Space Research Association, Columbia, MD, USA.
[24]Institute for Atmospheric Physics, Johannes Gutenberg University, Mainz, Germany.
[25]Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA.
[26]NASA Goddard Institute for Space Studies, New York, NY, USA.

Correspondence to: Julie M. Nicely (julie.m.nicely@nasa.gov)

Abstract. Hydroxyl radical (OH) plays critical roles within the troposphere, such as determining the lifetime of methane (CH_4), yet is challenging to model due to its fast cycling and dependence on a multitude of sources and sinks. As a result, the reasons for variations in OH and the resulting CH_4 lifetime (τ_{CH_4}), both between models and in time, are difficult to diagnose. We apply a neural network (NN) approach to address this issue within a group of models that participated in the Chemistry-Climate Model Initiative (CCMI). Analysis of the historical specified dynamics simulations performed for CCMI indicates that the primary drivers of τ_{CH_4} differences among ten models are the flux of UV light to the troposphere (indicated by the photolysis frequency JO^1D), mixing ratio of tropospheric ozone (O_3), the abundance of nitrogen oxides ($NO_x \equiv NO+NO_2$), and details of the various chemical mechanisms that drive OH. Water vapor, carbon monoxide (CO), the ratio of NO:NO_x, and formaldehyde (HCHO) explain moderate differences in τ_{CH_4}, while isoprene, CH_4, the photolysis frequency of NO_2 by visible light (JNO_2), overhead O_3 column, and temperature account for little-to-no model variation in τ_{CH_4}. We also apply the NNs to analysis of temporal trends in OH from 1980 to 2015. All models that participated in the specified dynamics historical simulation for CCMI demonstrate a decline in τ_{CH_4} during the analysed timeframe. The significant contributors to this trend, in order of importance, are tropospheric O_3, JO^1D, NO_x, and H_2O, with CO also causing substantial interannual variability in OH burden. Finally, the identified trends in τ_{CH_4} are compared to calculated trends in the tropospheric mean OH concentration from previous work, based on analysis of observations. The comparison reveals a robust result for the effect of rising water vapor on OH and τ_{CH_4}, imparting an increasing and decreasing trend of about 0.5 % decade^{-1}, respectively. The responses due to NO_x, O_3 column, and temperature are also in reasonably good agreement between the two studies.

Deleted: due mostly to clouds

Deleted: , though a discrepancy in the CH_4 response highlights a need for further examination of the CH_4 feedback on the abundance of OH…

1 Introduction

Hydroxyl radical (OH) is a key species of interest for numerous tropospheric chemistry studies over the past several decades. As a result of its role as the primary daytime oxidant in the lower atmosphere, OH determines how quickly many tropospheric gases and aerosols degrade or transform chemically. Notably, loss of atmospheric methane (CH_4) is dominated by its reaction with OH. Uncertainties in the abundance of OH at the global scale, coupled with source terms of CH_4 that are difficult to quantify, have driven disagreement in the causes of recent variations in the CH_4 growth rate (Nisbet et al., 2019; Turner et al., 2019). As a key element in the CH_4 budget, tropospheric OH must be studied further to clarify its present-day abundance as well as its variability over time.

Numerous studies have sought to constrain the OH abundance and resulting CH_4 lifetime (τ_{CH_4}) using observations, global atmospheric models, and combinations of the two. Historically, chemical inversion of methyl chloroform (MCF: CH_3CCl_3) comprised the primary method capable of gleaning information about global-scale OH burdens (Bousquet et al., 2005; Krol et al., 1998; Lovelock, 1977; Montzka et al., 2000; Prinn et al., 1987; Ravishankara and Albritton, 1995; Spivakovsky et al., 2000), though additional species that are lost by reaction with OH were also tested for this purpose (Jöckel et al., 2002; Liang et al., 2017; Miller et al., 1998; Nisbet et al., 2016, 2019; Singh, 1977; Weinstock and Niki, 1969). Models have likewise been relied upon to derive tropospheric OH abundance and its evolution. Stevenson et al. (2006) found a large spread in τ_{CH_4} (6.3 to 12.5 years) from a suite of atmospheric chemistry models in an analysis performed more than a decade ago. Seven years later, the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP) generated both historical (Naik et al., 2013) and future (Voulgarakis et al., 2013) simulations from numerous chemistry-climate models,

Deleted: Notably

revealing still-large discrepancies not only in present-day τ_{CH_4} (with values ranging from 7.1 to 14.0 years) but also in how τ_{CH_4} is expected to vary through year 2100 given common emissions scenarios. Note that, here and throughout, τ_{CH_4} refers to the lifetime of CH_4 due to reaction with tropospheric OH only. Most recently, the confluence of observations with advanced modelling techniques have enabled sophisticated analyses of global OH (Holmes et al., 2013; McNorton et al., 2016; Prather et al., 2012; Rigby et al., 2017; Turner et al., 2017). Despite the advent of numerous observing systems for species with some bearing on OH chemistry in the last several decades, it is widely acknowledged that current observations are insufficient to unambiguously derive current trends in OH (Nisbet et al., 2019; Prather & Holmes, 2017; Turner et al., 2019, 2017).

Deleted:

While global models are insufficient for clarifying the outstanding questions regarding OH and τ_{CH_4} on their own, they can serve as valuable testbeds in which to evaluate the factors influencing OH chemistry. The dominant reactions responsible for producing, cycling, and sequestering OH (see, e.g., Spivakovsky et al. (2000)) are well characterized and represented, to varying degrees of explicitness, in modern chemical mechanisms. Despite general consensus on the immediate drivers of OH chemistry, large differences in OH can manifest due to infrequently diagnosed differences in, e.g., ultraviolet (UV) flux to the troposphere (needed to initiate ozone (O_3) photolysis for subsequent OH primary production) due to variations in cloud parameterizations and radiative transfer codes. Similarly, differences in the representations of volatile organic compound (VOC) oxidation pathways can influence the extent to which OH is recycled following reactions with hydrocarbons. Such nuances in the chemistry of OH make OH differences between models notoriously difficult to attribute. With properly coordinated simulations and sufficient model output, however, we have demonstrated that the barriers posed by complex, non-linear chemistry can be overcome.

The multi-dimensional system that describes OH behaviour is well-suited for study via machine learning approaches. We have previously demonstrated the utility of neural networks (NNs) for quantifying differences in OH among a small group of chemical transport models (CTMs), which rely on the specification of meteorological conditions (Nicely et al., 2017). Other groups have similarly shown the promise of machine learning techniques to better parameterize within models such complex processes as convection (Gentine et al., 2018), radiative transfer (Krasnopolsky et al., 2009), ozone production (Nowack et al., 2018) and deposition (Silva et al., 2019), and to replace the numerical integrators that simulate chemistry within models (Keller and Evans, 2019). NNs in particular are capable of modelling complex non-linear functions, making them a suitable technique for studying the non-linear chemistry involved in OH production and loss. The community continues to develop best practices for harnessing the power of machine learning for applications in atmospheric science. We build here on the specific application of NNs to better understand model representations of OH.

In this study, we apply an NN approach to quantifying the causes of OH differences to the large group of models that participated in the Chemistry-Climate Model Initiative. We repeat our earlier analysis that identifies the primary drivers of

OH and τ_{CH_4} differences among model simulations conducted with specified dynamics, for a single year. We then expand the approach to study temporal variations in OH for 1980-2015, allowing for attribution of trends and interannual variability in τ_{CH_4} to specific parameters. Finally, we compare the derived trends in OH simulated by the CCMI models to trends derived from a previous observation-based study.

2 Model Simulations

The Chemistry-Climate Model Initiative (CCMI), carried out as an official activity of the International Global Atmospheric Chemistry (IGAC) and the Stratospheric Processes And their Role in Climate (SPARC) communities, seeks to enable inter-model evaluation of chemistry-climate models (Eyring et al., 2013). Phase 1 of CCMI has designed a set of simulations, covering both historical and future timeframes, with prescribed emissions inventories such that the interactive chemistry and its interplay with dynamical and radiative processes can be robustly compared between models. The analysis presented here focuses on one simulation, the historical specified dynamics (SD) simulation from 1980 to 2010 (REF-C1SD) (Hegglin and Lamarque, 2015; Morgenstern et al., 2017). Details of the emissions inventories recommended for this simulation can be found in Eyring et al. (2013). We have also performed the inter-model comparison portion of this analysis (Section 3.2) for the historical free-running simulation conducted from 1960 to 2010 (REF-C1). However, since a comprehensive examination of OH within the REF-C1 simulations was conducted by Zhao et al. (2019), those results are presented in the Supplement. We also include output from models that are not formal participants in CCMI, but provided simulations comparable to those being used here. These additional models are described below. Monthly mean fields are used for the various chemical, physical, and radiative parameters necessary for evaluating OH, described in Section 3. We analyse all models that include and provided output for the complete list of these variables.

Models that participated in the REF-C1SD simulation were nudged toward reanalysis meteorological fields such that dynamical conditions are represented with historical accuracy. The details of how nudging – of the winds, temperature, and sometimes pressure and water vapor fields – is conducted can be found in Morgenstern et al. (2017), Table S30. Particularly relevant to this analysis is the nudging of specific humidity, which is only performed in the MOCAGE model, of the models we analysed. Models that produced REF-C1SD simulations for CCMI and provided the necessary output to complete this analysis include: CAM4-Chem (Tilmes et al., 2016), EMAC-L47MA, EMAC-L90MA (Jöckel et al., 2016), MOCAGE (Guth et al., 2016; Josse et al., 2004), MRI-ESM1r1 (Deushi and Shibata, 2011; Yukimoto et al., 2012), and WACCM (Garcia et al., 2016; Marsh et al., 2013; Solomon et al., 2015). For both configurations of the EMAC model, the simulations that included nudging of wave-0 temperatures were used (Jöckel et al., 2016). All models, here and including those described below, include interactive stratospheric chemistry.

Deleted: further

Deleted:

Deleted: (

Deleted: ,

Deleted: Zhao et al. (2019)

Field Code Changed

Deleted:

Four models also contributed SD-type simulations to be analysed alongside the REF-C1SD CCMI simulations. The Goddard Earth Observing System (GEOS) model (Molod et al., 2015) conducted a “Replay” run, meaning the general circulation model computes its own meteorological fields for a 3 hour simulation period, then calculates the increment necessary to match a pre-existing reanalysis data set, in this case the Modern Era Retrospective Analysis for Research Applications version 2 (MERRA-2). The increment is then applied as a forcing to the meteorology at every time step during a second run of the same simulation period. This simulation includes full interactive tropospheric and stratospheric chemistry from the Goddard Modeling Initiative (GMI) chemical mechanism (Nielsen et al., 2017) with output for years 1980-2018 at 0.625° × 0.5° horizontal resolution and 72 vertical levels (Orbe et al., 2017; Stauffer et al., 2019; Wargan et al., 2018). This simulation is referred to as “GEOS Replay.” Additionally, three chemical transport models (CTMs), which directly rely on established meteorological fields such as MERRA-2 rather than calculate them, provided output used in this analysis. The OsloCTM and GEOS-Chem CTMs output all required variables for year 2000, while the GMI CTM (Strahan et al., 2013) simulated the full 1980-2015 period. All CTMs except GEOS-Chem calculate water vapor interactively in the troposphere. GEOS-Chem instead uses specific humidity fields from the MERRA reanalysis. We note that, while the GEOS Replay simulation described above used the GMI chemistry package, all discussion of the simulation from “GMI” refers to the separate, standalone CTM. While CTMs read in and use external meteorological fields rather than “nudging” or “replaying” internally calculated fields, we expect them to similarly represent realistic meteorological conditions for a given year. As such, we group them with the REF-C1SD simulations from CCMI, bringing the total number of SD-type simulations analysed to ten.

3 Methods

3.1 Neural network setup

Neural networks are generated to predict the monthly mean OH mixing ratio for a given model following the method outlined in Nicely et al. (2017). Briefly, four NNs are trained for one model, each for one simulation month. To reduce the computational demands of NN training, we only establish NNs for four months, one for each season: January, April, July, and October. Separate NNs are trained for the SD (main text) and free-running (Supplement) simulations, and all training is performed with output from year 2000. Each model gridbox located below the tropopause (thermal, following the WMO definition, for all models except GEOS Replay, which uses a “blended” tropopause calculation combining thermal and potential vorticity definitions) is a single sample, so sample sizes are determined by a model’s vertical and spatial native resolution. The number of tropospheric model grid points, and thus the training dataset sample size, is indicated for each model in Table S1 and always exceeds 100,000. Because separate NNs are trained for each month, and monthly mean output from each model simulation is used as input and training data, the dataset does not represent diurnal variations in OH chemistry.

Deleted: mimic the tropospheric chemistry with respect to OH

Deleted:

Deleted: one

Deleted: is

Deleted: at a time

The training process adjusts weighting factors such that mixing ratios of OH are predicted accurately when 3-D fields of the following variables are input to the NN: pressure, latitude, temperature (T), ozone (O_3), specific humidity (H_2O), methane (CH_4), the sum of nitrogen oxide and nitrogen dioxide ($NO_x \equiv NO + NO_2$), the ratio $NO:NO_2$, carbon monoxide (CO), isoprene (“ISOP”=C_5H_8), formaldehyde (HCHO), the photolysis frequency of NO_2 (“JNO_2”), the photolysis frequency of O_3 to excited state $O(^1D)$ (“JO^1D”), and stratospheric O_3 column (“O_3 COL”). Note that many of the inputs covary with one another depending on the chemical regime or meteorological conditions. A strength of the NN approach is that the inputs chosen need not be independent of each other. All chemical species are input to the NN as unitless mixing ratios, except for CH_4, which is normalized by the maximum tropospheric value and indicated by the notation “CH_4^{NORM}”. This normalization enables direct comparison of CH_4 distributions between models, despite the fact that the use of boundary conditions sometimes results in substantially different amounts of CH_4 between models. (While the CCMI models generally used roughly consistent boundary conditions, the additional simulations that were not formally part of CCMI exhibit CH_4 concentrations outside the ranges of those in the CCMI models.) Pressure is provided in units of hPa, temperature in K, photolysis frequencies in s^{-1}, and O_3 COL in Dobson Units (DU). Three of the inputs – HCHO, $NO:NO_2$, and O_3 COL – have been introduced to this analysis since Nicely et al. (2017), due to availability of output from all models and to the added information they encompass that may be relevant for OH chemistry. For instance, having knowledge of the partitioning of NO_x likely enables one to more accurately predict OH quantities compared to knowing just the total abundance of NO_x. Likewise, the introduction of O_3 COL is somewhat redundant when its primary effect on OH is through attenuation of ultraviolet (UV) flux to the troposphere, which is already encompassed by the input JO^1D. However, JO^1D is also altered by other factors such as clouds, which cannot as easily be included as an input for this analysis (some models provide 2-D cloud fraction fields, others output 3-D fields, and still others do not give any metric regarding clouds). Whether strong differences in JO^1D are caused by clouds or overhead O_3 should be clarified by inclusion of O_3 COL as an input.

Deleted: mutually exclusive

Deleted: to

Deleted: indication of their

Deleted:

Deleted: fields

The neural network architecture is consistent with that of Nicely et al. (2017) and is shown in **Figure 1**. However, the number of computational nodes was doubled from 15 to 30 given the availability of more powerful computing resources. Two hidden layers each containing 30 nodes provided strong performance of the NN in reproducing the OH mixing ratios from a given model. For training, the model output is randomly split 80%/10%/10% into training, validation, and test datasets. During that process, the data from the training set is used to actively adjust weighting factors, and the validation set is evaluated to determine a training stopping point. When errors in predicting the validation data grow after adjusting weighting factors some number of iterations in a row, it is determined that the NN model prior to the growth in errors likely reached a local minimum in its cost function. This manner of “early stopping” helps to prevent over-fitting, though application of the NNs to alternative years is not immune to over-fitting, an issue discussed further in Section 4.3.1. For further application of this method across varying time scales, we would recommend a more methodical approach to sampling model output in time as well as in space. The final 10% of data is then used to independently test the resulting NN, and compare between different training iterations. A total of five trainings were performed for each NN, and the NN with best

Deleted:

performance (evaluated by the correlation coefficient from comparison of NN-calculated and model-simulated OH values) was chosen as the NN to be used in further analysis. Further details of the training process and evaluation metrics can be found in Nicely et al. (2017).

Deleted: T

Deleted: are carried out as

Deleted: ; a detailed explanation of all procedures and methods may be found therein…

We note that alternative machine learning algorithms have seen increased application to problems within atmospheric science in the last few years, and may be equally or even better suited than neural networks to studying non-linear chemical systems. In particular random forest regressions and gradient boosting techniques offer greater computational efficiency and, in the case of random forests, have the capability to quickly identify which inputs are most strongly influencing the calculated output, known as "feature importance" (Grange et al., 2018; Hu et al., 2017; Keller and Evans, 2019; Liu et al., 2018). Additionally, linear regression algorithms such as Ridge and Lasso regression may be beneficial in curbing issues related to extrapolation. We also do not intend to suggest that our chosen NN input list, architecture, and general method is the best approach; input variables were largely determined by available output, and architecture testing was conducted on the computing resources available at the time of the study. It is possible that a single NN could suffice for predicting OH variations throughout an entire year, rather than for just a single month, following methodical subsampling methods to create the initial training dataset. As such, we encourage exploration of modifications to this method as well as additional algorithms for future machine learning applications to atmospheric chemistry.

3.2 Inter-model comparison approach

Once NNs are established for each model, an analysis is conducted to quantify the OH and τ_{CH_4} differences attributable to individual input terms. To accomplish this, each model, *A*, is paired with another model, *B*, such that one input to the NN of model *A* is substituted with the same field from model *B*. All other inputs are held fixed, using fields from model *A* for year 2000. Fields are interpolated to the resolution of the native model, *A* in this case, bilinearly across latitude and longitude, and linearly in log(pressure) space for the vertical coordinate. Any resulting changes in OH can then be directly attributed to the substituted variable.

The "swaps" that are performed in the manner described above undergo a process we refer to as "extrapolation control," which restricts the substituted variable from leaving the range of values over which the native model's NN was trained. If, e.g., O_3 is being substituted from CAM4-Chem into the GMI NN, we not only check that a given CAM4-Chem O_3 value lies within the minimum and maximum GMI tropospheric O_3 values, but also that the GMI value of CO at that gridpoint can be associated with the new CAM4-Chem O_3 value. This check is performed across all variables, and essentially prevents the substitutions from venturing too far outside of the chemical regimes simulated within the native model. In the case that a swapped variable exceeds the acceptable range of values, it is revised up or down accordingly. For reference, we tally the number of instances in which extrapolation control is invoked for two categories: coarse adjustments, when a NN input value from another model falls entirely outside the range of the NN input values from the native model, and fine adjustment, when

Deleted:

Deleted: is

a value from another model must be tweaked to preserve the native model's chemical regimes. On average, coarse adjustments are incurred for 3.5% of all swapped data points, while fine adjustments are made to 18.8% of swapped values. We find that extrapolation control is critical to achieve meaningful results with the NN inter-model comparison method, though it necessarily forces the attributed changes in OH and τ_{CH_4} to be conservative estimates.

Metrics used to evaluate the results of variable swaps include tropospheric OH integrated columns for visualization and changes in τ_{CH_4} for a globally-summed quantity. Tropospheric columns are integrated vertically and weighted by the mass of CH_4 and the temperature-dependent rate constant of reaction between OH and CH_4. The global mean lifetime of CH_4 is found using Eq. 1:

$$\tau_{CH_4} = \frac{\sum M_{air} \times \chi CH_4}{\sum [OH] \times k_{OH+CH_4} \times M_{air} \times \chi CH_4}, \qquad (1)$$

where M_{air} is the mass of air within a grid box, brackets denote number density, χ denotes mixing ratio, k_{OH+CH4} is the reaction rate constant for the OH + CH_4 reaction calculated for each grid box temperature, and summations are performed over all tropospheric model grid boxes. This formulation is equivalent to the standard lifetime calculation of burden divided by loss rate, adapted to the quantities most directly related to model outputs available (Chipperfield et al., 2014). Again, we note that this is strictly the atmospheric lifetime of CH_4 with respect to loss by tropospheric OH. If one additionally includes all stratospheric grid boxes within the above summation, annual average lifetimes of almost all models consistently increase by ~1.2 years.

Deleted:

Deleted:

3.3 Time series evaluation approach

A new element of this analysis applies the already-established NNs of each model to examine the time evolution of OH over several decades of simulation. For this, we focus on the REF-C1SD simulation set, as it contains the most realistic representation of historical emissions and meteorological conditions, and thus is most likely to resemble true OH variations. All models that provided SD-type simulations as described in Sections 2.2 and 2.3 are included, with the exception of GEOS-Chem and OsloCTM, both of which only provided output for year 2000. Using a similar swapping technique as described in Section 3.2, the NN for a given model is used to quantify the effect of substituting individual inputs from different years. No inter-model substitutions are conducted; instead, a single input is taken from the various years of the simulation (1980-2015) while all other inputs are fixed to their 2000 values. Because all swaps are performed on an intra-model basis, extrapolation control is largely unnecessary, since that model's chemical regimes do not vary drastically from the original year 2000 training output. However, we do see some instances, noted in Section 4.3, of anomalous behaviour in the τ_{CH_4} results because some variables undergo significant changes, particularly between the 1980s and training year 2000. Overall, the NN technique should be sufficiently generalizable to provide meaningful results even when using inputs lying modestly outside of the range of training values. Robustness of the results is demonstrated by the emergence of several consistent features between the eight models examined, as discussed in Section 4.

4 Results and Discussion

4.1 Native model and NN performance

Figure 2 shows values of τ_{CH_4} found for all models that produced SD-type simulations. Annually and globally averaged lifetimes vary from 6.59 years (OsloCTM) to 8.41 years (GMI). All models exhibit the expected seasonal variation in τ_{CH_4}, with minimum values in the Northern Hemisphere (NH) summer months due to higher OH at this time of year. Specifically, the seasonal variation in the global mean is a result of greater anthropogenic influence in the NH and resulting increases in concentration of two OH precursors: O_3 and NO_x.

An example of NN performance is shown for the January WACCM model in **Figure 3**, relative to the native model OH fields. Tropospheric OH columns are shown for the model and NN alongside the absolute value of the difference between the two. In general, the NNs from all models show similar magnitudes and spatial patterns in their calculated OH field, with errors somewhat randomly scattered and maximizing locally to values of ~10% of the total column value. Supplementary Figures S1-S4 show the performance of all NNs, for each of ten SD-type model simulations and for each of four months, while Table S2 provides further statistics on all NNs used here. Performance of all model NNs for year 2000 is strong, with values of τ_{CH_4}calculated from the NN-generated OH field within 0.006 years of the parent model's τ_{CH_4} on average. The maximum error in τ_{CH_4}, an overestimate by 0.012 years, occurs for the MRI-ESM1r1 model in the month of January. Performance is generally poorest in boreal winter, with average offsets in τ_{CH_4} of 0.007 years, and strongest in boreal summer, for which the mean bias is only 0.004 years.

4.2 Inter-model comparison

The inter-model comparison component of this analysis can be understood fundamentally by the OH and τ_{CH_4} differences generated by substituting input fields between models. An example of the OH column and τ_{CH_4} changes that are calculated through individual variable swaps is shown in **Figure 4**. The two models with the highest and lowest values of τ_{CH_4}, GMI and OsloCTM, respectively, are chosen for this example. Swaps performed between the two models for the month of January reveal that local O_3, JO^1D, HCHO, and NO_x account for the largest differences in τ_{CH_4} for this particular model pairing. A complete budgeting of the changes in τ_{CH_4} attributable to all inputs for GMI and OsloCTM is shown in **Table 1**. Note that the values of τ_{CH_4} shown in Table 1 correspond to lifetimes for the month of January rather than annual averages and so will differ from the lifetimes noted at the beginning of Section 4.1.

It is worth discussing several features that are evident in the visualized OH changes shown in Fig. 4. First is the spatial distribution of the OH variations. Depending on how the sink or source term undergoing the swap affects OH chemistry, the strongest impacts may occur in localized areas or may distribute evenly over the globe. For instance, varying local O_3 and NO_x (Fig. 4a, b and 4g, h, respectively) exert the greatest influence on OH over the climatological tropics, with maximum impacts over land but extending over the oceans as well. This is likely a result of the anthropogenic or biomass burning emissions sources, which limit the largest differences in O_3 and NO_x between the two models to areas proximate to the South American, African, and Indonesian source regions for the month of January. The OH changes resulting from substitutions of the inputs JO^1D and HCHO, however, are distributed over oceans as well as over land masses and, in the case of HCHO, are strongest in remote marine regions. This pattern is common for species that influence OH chemistry through mechanisms that are largely independent of local emissions. In the case of HCHO, its role as a secondary source of OH through CH_4 oxidation is relatively more important in the absence of large VOC concentrations, thus its stronger influence is seen away from terrestrial vegetation.

Deleted: continental

The second feature to note in Fig. 4 is the symmetry between input swaps in opposing directions. In other words, the swap of an input from OsloCTM into the GMI NN generally yields OH column and τ_{CH_4} changes that are equal but opposite to the changes resulting from use of a GMI input in the OsloCTM NN. With few exceptions, almost all regions of OH increase (red) in one model's NN are matched by OH decreases (blue) in the other model's NN in Fig. 4. The changes in τ_{CH_4} are correspondingly similar in magnitude but opposite in sign. This behaviour is expected because a swap that may, e.g., increase an OH precursor and subsequently cause an increase in OH for one model will manifest as a decrease in that same precursor when the substitution occurs in the NN of the other model. While this pattern occurs for the vast majority of cases across all model pairings and swaps performed for this analysis, there are instances when symmetry is not maintained. This could happen for two reasons. First, the sensitivities of the two models to a particular change in an OH precursor or sink could differ. For example, one model may be sensitive to an increase in isoprene, causing OH concentrations to drop in response. Another model may incorporate buffering effects (such as reactions involving oxidized volatile organic compounds (Lelieveld et al., 2016; Taraborrelli et al., 2012) that allow OH to be recycled following its reaction with isoprene, causing it to be less sensitive to the same change in CH_4. We refer to these variations in model sensitivities as chemical mechanism differences, as they are most likely a result of the chemical reactions, species representations, or reaction rates implemented within a model's chemical mechanism. The second explanation for lack of symmetry in the OH response to a model swap is a forced asymmetry in the swapped inputs themselves, imposed by the extrapolation control technique described in Section 3.2. It is possible that the swap of an input in one direction, i.e. from Model *A* into Model *B*, could proceed with no alteration to the substituted variable, while the swap in the other direction, i.e. from *B* to *A*, results in the variable lying outside the trained range of Model A. The extrapolation control process will revise the substitute variable field from Model *B*, such that the difference between it and the native field from Model *A* is lessened. As such, the first swap into the NN of Model *B* will yield a larger magnitude change in the input as compared to the swap into the NN of

Deleted: CH_4

Deleted: CH_4

Model *A*. The impact of these factors is indirectly quantified through a remainder term that falls out of a full budgeting analysis, described below.

A third consideration in interpreting the information presented in Fig. 4 is the conditions that must be met in order for a large change in OH to manifest through this analysis. First, the two models between which a swap is conducted must exhibit differences in the parameter of interest. Should the two models exhibit, e.g., very similar O_3 fields, then swapping one model's O_3 with the other's will produce little difference in the NN-calculated OH. Second, the model must have some OH sensitivity to the variable being swapped. If a model is insensitive to changes in CH_4, swapping in a drastically different CH_4 field may not cause a perceivable difference in OH. Therefore, the absence of an OH response does not necessarily mean that input fields are similar between models. Conversely, the existence of large OH changes indicates that differences in the swapped input field exist between the two models *and* that the native model demonstrates a dependence of OH on that input variable.

A fourth issue is the fact that NNs can exhibit some degree of random behaviour, based on how they were trained and initialized. Our method involved training 5 NNs and selecting from those the one that performed best when compared to the independent test dataset. That single NN was used in all subsequent analysis. However, it is a useful exercise to evaluate the role of NN randomness in our results. We show in Figures S5 and S6 the left and right panels of Fig. 4, reproduced for the alternate NN trainings of the GMI and OsloCTM models, respectively. A visual comparison of tropospheric OH column differences among the five trainings of each model's NN reveals markedly similar spatial distributions and magnitudes. The values of calculated $\Delta\tau_{CH_4}$ do differ somewhat between the training instance, with larger effects on some variable swaps than for others. For instance, the standard deviation of the values of $\Delta\tau_{CH_4}$ calculated for all five trainings of the GMI NN is about 0.2 years for the $J(O^1D)$ and HCHO swaps, but less than 0.05 years for O_3 and NO_x. We note, though, that some of the NNs displayed in Figures S5 and S6 exhibit worse performance than the one ultimately chosen for subsequent use. As a result of this exercise, the uncertainties resulting from this analysis method may be considered, at most, to be ~0.2 years.

The final point of interest in Fig. 4 is the general consistency in the signs of OH and τ_{CH_4} changes for each model. The substitutions of all four variables generally cause an increase in OH within the GMI NN (and corresponding decrease in τ_{CH_4}) and a decrease in OH (increase in τ_{CH_4}) within the OsloCTM. This feature is most pronounced for this particular pair of models due to our reasoning for choosing them: they exhibit the largest difference in τ_{CH_4} among our group of 10 models. Because the native GMI model has a longer τ_{CH_4} value compared to OsloCTM, it makes sense that incorporation of OsloCTM's various OH precursor and sink fields into the GMI NN will tend to decrease the GMI τ_{CH_4}, bringing it into closer agreement with that of OsloCTM. This characteristic points to the utility of this analysis as a budgeting tool for quantifying the cause of the difference in τ_{CH_4} between two models. The τ_{CH_4} accounting for the GMI and OsloCTM set of swaps conducted for January is shown in Table 1. When considering all 12 variable swaps that were performed, the NN analysis more than explains the original gap in τ_{CH_4} between the two models. The GMI January lifetime of 9.24 years is decreased to 6.71 years ($\tau_{ORIG} + \Delta\tau$) after summing all $\Delta\tau$ values, while the OsloCTM lifetime is increased from 7.18 years to 9.48. This budgeting rarely provides a perfect accounting of the τ_{CH_4} gap due to the same reasons that give rise to asymmetric OH responses to a given swap: chemical mechanism differences and asymmetric swaps of inputs due to extrapolation control. As a result, a remainder term, found as the difference between the other model's τ_{ORIG} and the present model's $\tau_{ORIG} + \Delta\tau$, is attributed to these factors. This term is listed in the last row of Table 1 with the label "Mech."

Results from analysing individual model pairs reveal a multitude of insights regarding idiosyncrasies in emissions of, global distributions of, and OH sensitivities to the various input parameters. These results, available at our FTP site provided in Data Availability, may be especially useful to the reader with an interest in a particular species or model. However, with over 4000 plots (12 species × 10 models × 9 sub models × 4 months = 4320) and 180 τ_{CH_4} budget tables generated, it is beyond the scope of this paper to highlight and explain every interesting feature. Instead, we aggregate the results across all models to identify some primary conclusions. **Figure 5** shows the change in τ_{CH_4} for a specific model and substituted input variable, averaged over all nine pairings. For example, the data point shown for CAM4-Chem JO^1D is calculated from the nine $\Delta\tau_{CH_4}$ values obtained when swapping the JO^1D fields from the other nine models into the CAM4-Chem NN. The circular point represents the mean of those nine values, while the whiskers indicate one standard deviation about the mean. Aggregate results shown in this manner are compiled both for individual months (available on the FTP site noted above) as well as for annually averaged output. The latter is calculated as the average of the four monthly mean and standard deviation values, and is shown in Fig. 5.

As with the individual OH tropospheric column change plots (Fig. 4), numerous conclusions can be drawn by studying the aggregated results in Fig. 5. The method for reading the data in Fig. 5 is demonstrated in the following example. The mean $\Delta\tau_{CH_4}$ value attributable to JO^1D for the WACCM model is +0.99 years. This indicates that use of JO^1D fields from other models causes τ_{CH_4} to increase by ~1 year, meaning the native JO^1D field from WACCM imparts a low bias to τ_{CH_4} of 1 year, relative to the other models. A low τ_{CH_4} would result from OH concentrations being too high. Since OH and JO^1D are positively correlated (i.e., JO^1D can be thought of as a source for OH) the too-high OH is an indication of too-high JO^1D. In general, positive values of $\Delta\tau_{CH_4}$ correspond to relative high biases in input parameters that are source terms for OH and to low biases for species that instead serve as sinks. This reasoning is less straightforward for species such as HCHO, which can both produce and consume OH, while it is also produced by OH-initiated oxidation. We stress that these comparisons are strictly relative to other models, not to any observation or other indication of truth. So, points that appear as outliers in Fig. 5 should not necessarily be interpreted as an erroneous result, but rather should be considered as an area for further examination.

The ordering of variables along the x-axis of Fig. 5 denotes the average magnitude of $\Delta\tau_{CH_4}$ values across all models, with parameters on the left accounting for the largest τ_{CH_4} differences. As such, JO^1D is the largest driver of OH differences in the CCMI SD model simulations, followed by local O_3 and NO_x. The subsequent variables H_2O, CO, the NO:NO_x ratio, and HCHO cause moderate variations in tropospheric OH, while ISOP, CH_4, JNO_2, O_3 COL, and T are not responsible for inter-model spread in τ_{CH_4}. We note that T differences between the SD simulations are likely limited due the meteorological constraints imposed on the models. However, examination of the free-running simulations, discussed in the Supplementary Material, also shows practically no impact of T on OH. Thus, we conclude that the effect of temperature on OH chemistry is likely indirect, acting through pathways embodied by other variables, such as H_2O and species that exhibit strongly temperature-dependent reaction rates. Finally, the Mech. term, described in the discussion of Table 1, appears on the far right, indicating its origins as a remainder term from the budget analysis of individual model pairs. The magnitudes of $\Delta\tau_{CH_4}$ values attributed to chemical mechanism differences and asymmetric swaps between models are large enough to consistently rank the Mech. term third, between O_3 and NO_x, in terms of importance for OH in this analysis. Especially in model simulations conducted with common emissions inventories, we expect some of the disparity in a short-lived species like OH to emerge from differences in chemical mechanism implementations. In other words, when responses in OH to a given change in a source or sink term differ between two models, the remainder term (or term labelled "Mech." in Table 1) will increase, representing variations in the sensitivity of OH that presumably arise due to the two different implementations of the chemical mechanism. It is possible that other factors are represented by this term; e.g., other chemical species that influence OH chemistry but are not considered in the NN analysis could contribute to the Mech. term. However, previous analysis using a 0-D chemical box model as a "standard" mechanism in Nicely et al. (2017) suggested a correlation between actual biases in OH imparted by a given model's chemical mechanism and the remainder term resulting from the NN analysis. Therefore, we have some confidence that the Mech. term is meaningful, though significant further study would be required to parse the actual mechanistic differences responsible for imparting bias in OH calculations.

Significant inter-model differences in the largest driver of τ_{CH_4} spread, JO^1D, could arise from two possible sources. The amount of solar UV light penetrating down to the troposphere is largely dictated by the stratospheric column O_3 amount. However, the differences in total O_3 column are generally small and insufficient to cause the variations in JO^1D seen among the CCMI models. Rather, JO^1D likely varies to a great extent due to differences in cloud cover, and dissimilar treatments of clouds within model photolysis codes. Figure S7 highlights this effect by showing the ratio of JO^1D at the surface to JO^1D in the upper troposphere (UT) for each model. The relatively small column amounts of O_3 within the troposphere should account for very little absorbed UV light, making it much more likely that deviations in this ratio from 1.0 are driven by scattering due to clouds and possibly aerosols. The fact that models show large spatial differences in this ratio is a strong indication that clouds underlie the model differences in JO^1D.

Deleted: Supplementary

Deleted: 1

While the model differences in JO^1D, O_3, NO_x, and chemical mechanisms appear to drive the bulk of the τ_{CH_4} spread among this group of CCMI models, we emphasize that individual models may not adhere to these conclusions. As such, any efforts to improve a particular model should instead focus on the results specific to that model. For instance, HCHO plays a very small role in describing inter-model differences in OH on average, but for the OsloCTM model, HCHO is a much more important factor. Thus, we refrain from offering an across-the-board solution for remedying the large model spread in τ_{CH_4} and instead suggest a more individualized approach of studying plots such as those shown in Fig. 4 for more spatially and temporally resolved information. Visualizations of all model swaps, for all months and species, are available at our FTP site provided in Data Availability for this purpose.

There are several other qualifications to note when considering the results of the inter-model comparison. One is the negating effect between the JO^1D and tropospheric O_3 variables. Many, but not all, model $\Delta\tau_{CH_4}$ values for JO^1D in Fig. 5 are opposite in sign to the $\Delta\tau_{CH_4}$ values attributed to O_3. Physically, photolysis of tropospheric O_3 by light at wavelengths below 336 nm to form excited state $O(^1D)$ and subsequent reaction with H_2O to form OH is a loss pathway for O_3. Therefore, more UV flux will tend to decrease tropospheric O_3 concentrations while increasing OH, and vice versa. This physical mechanism, then, can explain the frequent cancellation of the $\Delta\tau_{CH_4}$ values attributed to these two factors. Should a modeler attempt to alter a model's OH field by forcing adjustments in its JO^1D, the opposing impact of tropospheric O_3 may result in no change for the value of τ_{CH_4}. However, this does not preclude the finding that both JO^1D and tropospheric O_3 are substantially different in the models for reasons we do not fully understand. Tropospheric O_3 can also vary between models for reasons external to the radiative environment. For instance, differences in the stratosphere-troposphere exchange, wet and dry deposition, and lightning NO_x emissions can each cause substantial variations in tropospheric O_3 among models (Wild, 2007). Further parsing of the reasons for the O_3 differences seen among the CCMI models is difficult without specialized output, including tracers such as ozone of stratospheric origin and NO_x generated by lightning. We recommend a targeted study to address the underlying reasons for the variations in tropospheric O_3.

Another qualification concerns the issue of causation versus correlation. Machine learning techniques, and NNs in particular, are generally more adept at identifying the predictors of a certain phenomenon than traditional methods, such as multiple linear regression. However, it is still possible that an input that is tightly correlated with the output may be misidentified as a driver of variations in the output. This is particularly relevant to keep in mind for species that serve as sinks of OH, such as CO and CH_4. Whether a decline in OH initiates or results from an increase in its sinks is difficult to differentiate, even with advanced analysis methods. Therefore, descriptions of CO and CH_4 as drivers of OH variations in this text may just as well be interpreted conversely, as downstream indicators of the change in oxidizing capacity.

Deleted: The

A final qualification is this analysis constitutes a foundationally hypothetical experiment. It essentially addresses the questions, "What if we could switch the fields of just one chemical species between two global models? What would be the instantaneous impact on OH? on τ_{CH_4}?" This approach, then, necessarily neglects the roles of feedbacks in the atmospheric system (e.g., if the NO_x field is perturbed, this will propagate to changes in O_3 as well, with time). However, for the objective of teasing apart the influences on global OH abundance and τ_{CH_4} and explaining inter-model differences, a notoriously difficult task, we regard our approach as a valuable exercise.

4.3 Time series evaluation

The second half of our NN analysis interrogates temporal trends in OH and τ_{CH_4}. **Figure 6** shows the evolution of τ_{CH_4} in the SD-type simulations conducted for 1980-2010. Two models, GEOS-Chem and OsloCTM, only provided output for year 2000, and so only appear as single points in Fig. 6. In addition, some models provided output beyond year 2010; output from years through the end of 2015 was included when available. The lifetimes all show a general downward trend over time, consistent with the upward trend in global mean tropospheric OH concentration shown by Zhao et al. (2019b) (their figure 4). Results concerning attribution of the τ_{CH_4} time series are presented in subsection 4.3.1, while derivation and analysis of trends are shown in subsection 4.3.2.

4.3.1 Attribution of the τ_{CH_4} time series

Swaps of input variables to a NN are conducted on an intra-model basis, with the goal of determining which OH precursors and sinks are responsible for OH variations over time. The results of these swaps are shown for each model in **Figure 7**. Changes in τ_{CH_4} attributable to each parameter are displayed as a function of year. Because we use the same NNs established for the inter-model comparison described in Section 3.2 trained on output from year 2000, the values of $\Delta\tau_{CH_4}$ for all species in year 2000 of Fig. 7 is zero by design. As an input field from another year is swapped into the NN, however, OH differences manifest and are denoted by the corresponding change in τ_{CH_4}. Because we are relying on the same NNs used for the inter-model analysis, we emphasize that the CH_4 fields used here are still normalized, separately for each year. As a result, the variations in τ_{CH_4} due to CH_4 should not be interpreted as a measure of the CH_4 feedback factor (Fiore et al., 2009; Holmes, 2018; Holmes et al., 2013; Prather et al., 2001). Instead of representing the change in OH with a change in absolute concentration of CH_4, the numbers shown here signify the change in OH with a change in how CH_4 is distributed within the atmosphere, both vertically and spatially. Largely, one would expect this to remain constant over time, though results from this analysis of the CCMI simulations suggests there are some modest changes in τ_{CH_4} attributed to the distribution of tropospheric CH_4. Should a similar method be applied to analysis of temporal variations in OH in the future, we would encourage training the machine learning algorithm on data spanning all years such that use of CH_4 absolute values would be possible.

While significant diversity in the drivers of OH variability across models is evident from Fig. 7, there are also several distinctive features that appear repeatedly. For instance the response of τ_{CH_4} to changes in CO shows a prominent peak in

Deleted: (

Deleted: ,

Deleted: Zhao et al. (2019)

Formatted: Heading 2, Space After: 0 pt year 1998 in all models except one. To gauge the role of emissions in this response, we show in Supplemental Figures S8-12 the time series of CO mixing ratios and other parameters averaged for the region most impactful to τ_{CH_4}: the tropical lower troposphere (latitudes between 30°S and 30°N, pressures greater than or equal to 700 hPa). Indeed, CO mixing ratios maximize in almost all models in year 1998, likely as a result of the emissions inventory reflecting the extreme biomass burning and strong El Niño Southern Oscillation (ENSO) event during that and the preceding year (Duncan et al., 2003 and references therein). The increase in τ_{CH_4} can thus be explained by the increased CO sink of OH, causing a temporary depletion of the oxidant. In addition, less distinctive peaks in τ_{CH_4} due to CO are identified in other years with strong El Niño conditions, notably 1982-1983, 1987, and 1991-1992 (Duncan et al., 2003).

The impacts of several other variables on τ_{CH_4} also demonstrate behaviour with reasonably identifiable causes. A prolonged decrease in τ_{CH_4} due to JO^{1}D from 1992 to 1998 is evident in the analysis of the CAM4-Chem, GEOS Replay, GMI, MRI-ESM1r1, and WACCM NNs. This may correspond to several confounding events that acted to increase the flux of UV light to the troposphere, increasing the primary production of OH and decreasing τ_{CH_4}, as seen in Fig. 7. First, solar activity reached a maximum around 1990, after which the decline in sunspots correlated strongly with a decline in tropical total O_3 columns (Duncan and Logan, 2008). Second, the eruption of Mount Pinatubo in 1991 likely impacted JO^{1}D through the decrease in stratospheric O_3 that resulted (Aquila et al., 2013; Tie and Brasseur, 1995). Finally, the prolonged ENSO event of 1990-1995 (Allan & D'arrigo, 1999) may have caused reduction in cloud cover due to drought conditions (Duncan et al., 2003). Interestingly, the τ_{CH_4} response to H_2O is moderately anticorrelated with CO. This is particularly evident for year 1998 in many of the models, when large biomass burning events occurred in many regions of the world, such as the boreal forests of both Asia and North America, Central America and Mexico, and Indonesia, which were attributed in part to a strong El Niño in 1997 that transitioned in a strong La Niña in 1998. Although strong ENSO events cause drought conditions over some regions, it is more fundamentally associated with warming sea surface temperatures and increased evaporation, particularly in the tropical Pacific Ocean. Thus, it is reasonable that larger values of specific humidity will tend to increase OH primary production during an El Niño year, as suggested by the decrease in τ_{CH_4} shown in Fig. 7. An apparent increase in O_3 also coincides with the 1998 ENSO event, determined by the decreasing component of τ_{CH_4}. The prevalence of biomass burning would indeed cause increases in tropospheric O_3 through increased emissions of its precursors, CO, VOCs, and NO_x. Additionally, the τ_{CH_4} response to O_3 shows the most distinguishable trend of all the variables over the full 1980-2015 period. Steady decreases in τ_{CH_4} due to O_3 imply an increasing tropospheric O_3 burden, a modelling result supported by observations (Verstraeten et al., 2015).

We also note the appearance of spurious results in several cases. The τ_{CH_4} responses to CH_4 in EMAC-L47MA and EMAC-L90MA as well as to O_3 COL in MOCAGE extend to very large negative values in the early part of the time series. To show the full extent of the EMAC τ_{CH_4} responses to CH_4, we show alternate versions of Figs. 7b and 7c with expanded y-axis

Deleted: 2

Deleted: 6

Deleted:

Moved (insertion) [2]

Deleted: .

Deleted: s

Deleted: with both

Deleted: and

Field Code Changed

Deleted: ; Duncan et al., 2003

Deleted: . The former event likely impacted JO^{1}D through the decrease in stratospheric O_3 that resulted (Aquila et al., 2013; Tie and Brasseur, 1995), while the latter …

Deleted: Both effects will tend to increase the flux of UV light to the troposphere, increasing the primary production of OH and

Moved up [2]: and decreasing τ_{CH_4}, as seen in Fig. 7.

Deleted:

Deleted: the

Deleted: mentioned above was driven specifically by El Niño conditions.

Deleted: a

Deleted: El Niño

Deleted: s

Formatted: Font color: Text 1, English (US) ranges in Supplementary Figure S13. Chemical conditions during the 1980s would differ most markedly from the regimes simulated in year 2000, on which the NNs are based. Particularly for concentrations of CH_4, which underwent monotonic rise aside from a stabilisation period from 2000 to 2007 (Turner et al., 2019), conditions in 1980 could be quite different. However, as was noted in Section 3.1, CH_4 inputs to the NNs are normalized against the maximum tropospheric value. The field of CH_4 for each year is likewise normalized against the maximum CH_4 for that year, so a strong response in τ_{CH_4} must indicate a significant change in the distribution of CH_4, not just in changes in its concentration over time. Indeed, Supplementary Figure S14 shows the normalized CH_4 values used as input to the NNs for the pressure level closest to the surface. For each EMAC configuration (for the month in which the τ_{CH_4} response shown in Fig. 7 is largest and most unphysical), the CH_4 distributions in the 1980s do show notable change from the year 2000 distribution used for training. Specifically, relative CH_4 mixing ratios in the Southern Hemisphere drop relative to the larger concentrations in the Northern Hemisphere. Other models, such as WACCM shown in the bottom panels of Fig. S14 show practically no inter-annual change in the CH_4 distribution for a given month. This behaviour in the EMAC model likely results from implementation of a Newtonian relaxation scheme to determine a time-varying, latitude-dependent lower boundary condition for CH_4 (Jöckel et al., 2016). Our spurious NN result may indeed be explained by a slowdown in the rate of increase in CH_4 concentrations at the lower boundary initiated in 1980, evident in supplementary figure E1 of Jöckel et al. (2016). While this method of determining boundary conditions generally represents a more sophisticated treatment of CH_4, within the context of this analysis, it imparts an artificially strong signal in OH and τ_{CH_4}. Therefore, the unphysical results in Fig. 7b and 7c due to CH_4 indicate an artefact due to the NN method, not a problem in the EMAC model itself.

Deleted: 7

Deleted: 7

Deleted: appears to

Deleted:

For the other occurrence of anomalous behaviour, MOCAGE shows an unrealistically large response of τ_{CH_4} to O_3 COL in the 1980s (Fig. 7f), a result not corroborated by any other model. Supplementary Figure S15 illustrates the likely cause of this behaviour. While most models exhibit modest changes in total O_3 COL between 1980 and 2000, including GEOS Replay shown in the top set of panels, the MOCAGE model (bottom panels) shows much larger column amounts in year 1980. These values fall well outside the range of O_3 COL amounts on which the NN was trained, so unrealistic behaviour of the NN in this case is not surprising.

Deleted: 8

These examples of spurious results highlight an issue that must be treated with caution when using machine learning approaches. Because the application of our NN method to time series analysis is an extension beyond the originally intended purpose, not all NNs are sufficiently generalizable to reliably reproduce OH for years other than the training year, 2000. To account for this, we evaluate each NN for all years by inputting variables from each year. With this test, all inputs are changed, not just a single input at a time. The resulting OH, as depicted in Figures S16-S23 for select years, compares well to the native model's OH field for that year in many cases, but not in all. Considerable bias occurs at low OH mixing ratios, though we note that near-zero concentrations will likely not affect the resulting globally-integrated τ_{CH_4} unless values are grossly overestimated. This evaluation also represents a rigorous test of the NNs, as significant shifts in numerous inputs at

Deleted:

once might push the NN algorithm into new phase space not encountered during training, much more so than only changing one input at a time, which is our approach in the subsequent time-series analysis. Nonetheless, we limit the influence of poorly generalizable, or "overfit," NNs by only including in the multi-model mean results for the years in which a NN reproduces its native model's OH field with an r^2 value greater than or equal to 0.95. For four NNs (one per month) created for each of 8 CCMI models, across 36 years, the potential application of the NNs to 1152 calculations (4×8×36) is reduced to 696 calculations using this test. Results from this point forward are subject to this quality check, and were found to be insensitive to the r^2 threshold imposed. This insensitivity is demonstrated by alternate versions of the figures to come, placed in Supplement, generated using all NNs rather than the quality-filtered NNs.

Deleted: Due to the anomalous NN behaviour in these three cases, the responses to EMAC CH_4 and MOCAGE O_3 COL are excluded from subsequent analysis of the NN time series results.¶

Figure 8 shows the multi-model mean attribution of variations in τ_{CH_4}. Many of the same features identified in Fig. 7 also emerge here: clear definition of strong ENSO years in the CO response, apparent Mt. Pinatubo effects in the JO^1D response, and a general downward trend in τ_{CH_4} due to O_3 are all observed. Also, as might be expected from the inter-model comparison results discussed in the prior section, JO^1D, O_3, NO_x, H_2O, and CO account for many of the strongest OH variations over time (Fig. 7) as well as between models (Fig. 5). Supplementary Figure S24 shows the analogue of Fig. 8, without the quality filter applied to the NNs described above. I.e., all NN results from Fig. 7 are included, except the spurious cases of EMAC CH_4 and MOCAGE O_3 COL.

4.3.2 Trends and interannual variability in the τ_{CH_4} time series

Formatted: Heading 2, Space After: 0 pt

We also perform linear fits to each response time series in Fig. 8. The resulting trends in τ_{CH_4} are shown in **Figure 9**, panel (a). The interannual variability of τ_{CH_4} is also calculated as the standard deviation of the detrended time series, shown in Fig. 9b, though it is relevant to note that CTMs have historically not captured the full interannual variability exhibited by observed OH proxies (Holmes et al., 2013). Supplementary Figure S25 shows the equivalent of Fig. 9, without application of the NN quality filter described above. Negative trends in τ_{CH_4} due to O_3, H_2O, JO^1D, and NO_x stand out as largest in magnitude. The sum of all factors shown in Fig. 9a is –2.3±0.4% decade^{-1}, which is comparable to the mean downward trend in τ_{CH_4} seen in Fig. 6, –1.8% decade^{-1}. Time series of the model input variable fields show corresponding trends, with parameters that serve as source terms of OH increasing over time (Supplemental Figures S9-12). Tropospheric O_3 and NO_x show clear upward trends over time, while H_2O and JO^1D show upward trends with more variability, which is also conveyed by the error bars in Fig. 9a. It is interesting to note that H_2O plays a stronger role in the overall temporal trend of τ_{CH_4}, as compared to its role in explaining inter-model differences. This is likely due to the fact that temperatures were constrained in the specified dynamics simulations, which in turn should determine the water vapor calculated within the models. The interannual variability attributed to CO in Fig. 9b is also consistent with the large year-to-year swings in tropical lower tropospheric CO mixing ratios shown in Supplemental Figure S8. While Fig. 9a suggests that CO exhibits very little overall trend between 1980 and 2015, we note there is a discernible increase in CO prior to ~1998 in Fig. S8 followed by a steady

Deleted: ,

Deleted: and O_3

Deleted: 1.9

Deleted: 2

Deleted: 5

Deleted: 2

Deleted: 2

decline thereafter. This is consistent with remote site measurements that show significant negative trends in CO since the late 1990s (Zeng et al., 2012).

Finally, the attributed trends in τ_{CH_4} from the CCMI models (Fig. 9a) are compared in **Figure 10** to trends in tropospheric mean OH concentration ("$[OH]^{TROP}$") from a previous observation-based analysis (Nicely et al., 2018). In that work, TOMS/OMI/SBUV observations of total column O_3 were used to infer radiative effects on the OH burden, while water vapor from the AIRS instrument, CH_4 from surface observations, NO_x from a global model simulation constrained to realistic emissions, and temperature from the MERRA-2 reanalysis were analysed to calculate chemical impacts on $[OH]^{TROP}$. In Nicely et al. (2018), the trend in $[OH]^{TROP}$ due to NO_x encompassed the effects of both the total abundance and the partitioning of NO_x, while the O_3 COL factor encompassed all radiative effects on OH. Thus, to perform a "like for like" comparison, the τ_{CH_4} trends due to NO_x and $NO:NO_x$ are combined, as are the trends due to O_3 COL and JO^1D shown in Fig. 9a. Error bars shown in Fig. 10 represent the 1σ uncertainty in the slope of the linear fit and, in the case of combined trends, are found by summing in quadrature the individual uncertainties. Because τ_{CH_4} varies with the inverse of OH concentration, note that the x-axis of Fig. 10 is inverted and a –1:1 line is shown in grey.

The trends in τ_{CH_4} from this analysis and in $[OH]^{TROP}$ from Nicely et al. (2018) are in reasonably good agreement for H_2O, NO_x, O_3 COL, and temperature. In particular, the two trends due to H_2O agree within the uncertainties, with τ_{CH_4} decreasing by ~0.5 % decade^{-1} and $[OH]^{TROP}$ increasing at almost the same rate. The impacts of NO_x and O_3 COL are found to increase OH concentrations in both studies, though the impacts on τ_{CH_4} from the CCMI models are found to be larger in magnitude than the observational estimate. The small impact of temperature, tending to lessen the OH burden, is also in close agreement between the two studies, with the CCMI models again showing a slightly stronger response. The role of NO_x in driving ~0.3 % decade^{-1} decline in τ_{CH_4} is roughly consistent as well. Only the effect of O_3 column falls relatively far from the –1:1 line, with analysis of the CCMI models suggesting a stronger decrease in τ_{CH_4} between 1980 and 2015, albeit with large uncertainties. This may result from inaccurate representations of stratospheric O_3 in the CCMI models, mischaracterization of the impacts on UV photolysis in the troposphere, or a combination of both. Overall, the results depicted in Fig. 10 show relatively robust findings regarding the responses of $[OH]^{TROP}$ and τ_{CH_4} to the factors examined through these two independent studies.

Because the CH_4 used as input for the CCMI NNs was normalized, as discussed above, the trend in τ_{CH_4} found in this analysis due to CH_4 did not represent a CH_4 feedback factor in the traditional sense. As such, it is not comparable to the trend in $[OH]^{TROP}$ due to CH_4 found by Nicely et al. (2018) and so was not included in Figure 10. However, even in the event that one were to retrain new NNs using absolute values of CH_4 and sampling across all years to generate the training dataset, we would question the physical meaning of the resulting trends. With the current necessity of providing boundary conditions for surface CH_4 rather than fluxes in models, our ability to realistically simulate CH_4 is hampered. We encourage

Moved (insertion) [1]

Deleted: t

Deleted: The response of OH to CH_4 however, is in poor agreement between the two studies. The previously determined observation-based estimate of $[OH]^{TROP}$ trend due to CH_4 was –1.01±0.05 % decade^{-1} while the CCMI model-based trend in τ_{CH_4} is only +0.06±0.07 % decade^{-1}. On one hand, the treatment of CH_4 as a normalized value within the NN analysis, as noted above in the discussion of Fig. 7, precludes a realistic estimate of the OH response to changes in CH_4. Rather, the trend estimate calculated by the NN analysis of CCMI models represents the impact on OH of changes in the distribution of CH_4 within the troposphere. Since the source regions of CH_4 are not expected to change substantially over the 1980-2015 period, it is not surprising that the CCMI model-based trend is small. Meanwhile, the other estimate of the $[OH]^{TROP}$ trend due to CH_4 from Nicely et al. (2018) is not without limitations. As was acknowledged in that paper, the box model method used to estimate the sensitivities of OH to CH_4 (among other species) is inherently inadequate for capturing complex coupling of chemical systems and downstream effects. For example, the box modelled sensitivity of OH to variations in CH_4 were found for a range of latitude, pressure, and NO_x values (since the latter determines whether CH_4 oxidation consumes or regenerates OH radicals). To maintain the characteristics of those chemical regimes, then, O_3 was input and held fixed in the box model simulation. As a result, especially in the relatively low-NO_x conditions prevalent throughout much of the troposphere, an increase in CH_4 would tend to consume OH without the corresponding increase in O_3 expected to result from greater CH_4 oxidation. That increase in tropospheric O_3 would offset some of the OH loss by increasing primary production, a process that should be captured in a fully coupled chemistry-climate model like those participating in CCMI.¶

the further examination of the response of OH to CH_4 on the global scale, which is likely a large influencer of tropospheric OH abundance, as indicated in Nicely et al. (2018) and Holmes et al. (2013).

Deleted: ,

Deleted: known as the CH_4 feedback,

Moved up [1]: the results depicted in Fig. 10 show relatively robust findings regarding the responses of $[OH]^{TROP}$ and τ_{CH_4} to the factors examined through two independent studies.

Deleted: This topic has been of interest for some time (Holmes et al., 2013; Prather et al., 2001), though the necessity of providing boundary conditions for surface CH_4 rather than fluxes in models hampers our ability to realistically simulate CH_4. Regardless, a model approach using fully-coupled tropospheric chemistry, such as that performed by Holmes et al. (2013) for three CTMs, would provide a more direct measure of the CH_4 feedback on OH than both approaches depicted in Fig. 10. Except for trends attributed to CH_4,

5 Conclusions

We perform a neural network analysis of the monthly mean output from historical simulations of ten models that participated in CCMI for the purposes of understanding OH and τ_{CH_4} differences and temporal trends. NNs are trained to reproduce OH mixing ratios for a given model using 3-D fields of 12 OH precursor and sink parameters. Performing swaps of the NN inputs between models produces a quantitative estimate of the difference in τ_{CH_4} that can be attributed to variations in the substituted variable. Among the ten models that we examine, on average, variations in JO^1D, local O_3, NO_x, and chemical mechanisms account for the largest differences in τ_{CH_4}. Model diversity in representations of H_2O, CO, the partitioning of NO_x, and HCHO is responsible for moderate OH differences, while isoprene, CH_4, JNO_2, overhead O_3 column, and temperature account for little-to-no variation in OH. However, the relative importance of a particular variable is highly model-dependent, so any effort to improve the representation of OH within a given model should be guided by that particular model's results.

We also analyse time series of τ_{CH_4} using the year 2000 NNs generated for the first half of the study. All models exhibit a downward trend in τ_{CH_4} between 1980 and 2015, varying from –0.54 % decade^{-1} to –2.97 % decade^{-1} (average of –1.83 % decade^{-1}). Swaps of NN inputs are conducted between years rather than between models, so attributions of the factors influencing trends in τ_{CH_4} are found for each model and then combined into a multi-model mean result. This analysis indicates that the largest contributors to the decreasing trend in τ_{CH_4} are O_3, JO^1D, NO_x, and H_2O, while CO also imparts a large degree of interannual variability. Features due to strong ENSO events and associated biomass burning as well as the eruption of Mount Pinatubo are discernible in the time series of attributed variations in τ_{CH_4}. In particular, the species CO, H_2O, and O_3 instigate prominent responses during strong El Niño years. Finally, the attributed trends in τ_{CH_4} from the NN analysis of CCMI model output are compared to trends in tropospheric mean OH concentration found previously in the observation-based study of Nicely et al. (2018). While the strong response of τ_{CH_4} to increasing H_2O over time appears to be a robust result, disagreement on the CH_4 feedback on OH between the two studies highlights limitations in the approaches of both, in addition to more systemic issues in the community's ability to model CH_4.

The NN and machine learning methods in general provide a valuable tool for performing insightful model intercomparisons of complex systems in a computationally-efficient manner. These approaches, however, must be undertaken with care to avoid erroneous results and recognition of their limitations. At present, we have devised a method to identify the drivers of OH variations, whether between models or between years, at coarse temporal resolution. Much future work is needed, though: observations must be incorporated to introduce a ground truth element to this analysis in a manner that either adjusts for or avoids disconnects between coarse versus local/instantaneous spatiotemporal scales and appropriately accounts for measurement uncertainty; an analysis of model output with much higher temporal frequency is needed to identify exactly where model differences in chemical mechanisms lie; and subsequent studies of why the various OH precursor and sink fields differ are required to make this analysis of greatest utility for improving model representations of τ_{CH_4}. While these challenges are significant, they are not insurmountable, especially as machine learning and other advanced statistical analysis techniques continue to be developed and honed.

Data Availability

All output from most of the models that participated in CCMI is available at the Centre for Environmental Data Analysis (CEDA), the Natural Environment Research Council's Data Repository for Atmospheric Science and Earth Observation, at http://data.ceda.ac.uk/badc/wcrp-ccmi/data/CCMI-1/output. WACCM and CAM4-Chem output for CCMI is available for download at http://www.earthsystemgrid.org. For instructions for access to both archives see http://blogs.reading.ac.uk/ccmi/badc-data-access. Output from the models that were not formal participants in CCMI Phase 1 is available from the co-authors who performed the model simulations; please contact the corresponding author with requests. A complete set of figures and tables generated by the model intercomparison and time series analyses is available at the FTP site https://acd-ext.gsfc.nasa.gov/anonftp/acd/atmos/jnicely/. In the event that this site is no longer active, please contact the corresponding author for access to all results.

Acknowledgments

JMN was supported by an appointment to the NASA Postdoctoral Program at the NASA Goddard Space Flight Center, administered by the Universities Space Research Association under contract with NASA. The authors also acknowledge the joint WCRP SPARC–IGAC Chemistry-Climate Model Initiative (CCMI) for organizing and making available the suite of model simulations used here. Special thanks is extended to the modeling groups which, at times, provided extra output that enabled this inter-comparison to take place with the maximum number of participants. We also thank many colleagues who engaged in helpful discussions that shaped the direction of this work and inspired additional analyses, including, but not limited to: Vaishali Naik, Sarah Strode, Jason St. Clair, and Melanie Follette-Cook. This work was partly supported by JSPS KAKENHI Grant Number JP19K12312. LER acknowledges China Southern for partial support. The EMAC simulations have been performed at the German Climate Computing Centre (DKRZ) through support from the Bundesministerium für Bildung und Forschung (BMBF). DKRZ and its scientific steering committee are gratefully acknowledged for providing the HPC and data archiving resources for this consortial project ESCiMo (Earth System Chemistry integrated Modelling). GZ and OM acknowledge funding under the New Zealand Government's Strategic Science Investment Fund (SSIF).

Author contributions

JMN and RJS conducted initial design of the method. JMN carried out the analysis. Development and refinement of the analysis were further guided by BND, GMW, and TFH. All other authors provided model output central to the analysis. JMN drafted the manuscript, and all co-authors provided assistance in finalizing the figures and text.

Competing interests

The authors declare that they have no conflicts of interest.

References

Allan, R. J. and D'arrigo, R. D.: "Persistent" ENSO sequences: how unusual was the 1990-1995 El Niño? [online] Available from: https://search.proquest.com/docview/220515300?pq-origsite=360link (Accessed 22 July 2019), 1999.

Aquila, V., Oman, L. D., Stolarski, R., Douglass, A. R., Newman, P. A., Aquila, V., Oman, L. D., Stolarski, R., Douglass, A. R. and Newman, P. A.: The Response of Ozone and Nitrogen Dioxide to the Eruption of Mt. Pinatubo at Southern and Northern Midlatitudes, J. Atmos. Sci., 70(3), 894–900, doi:10.1175/JAS-D-12-0143.1, 2013.

Bousquet, P., Hauglustaine, D. A., Peylin, P., Carouge, C. and Ciais, P.: Two decades of OH variability as inferred by an inversion of atmospheric transport and chemistry of methyl chloroform, Atmos. Chem. Phys., doi:10.5194/acp-5-2635-2005, 2005.

Chipperfield, M. P., Liang, Q., Strahan, S. E., Morgenstern, O., Dhomse, S. S., Abraham, N. L., Archibald, A. T., Bekki, S., Braesicke, P., Di Genova, G., Fleming, E. L., Hardiman, S. C., Iachetti, D., Jackman, C. H., Kinnison, D. E., Marchand, M., Pitari, G., Pyle, J. A., Rozanov, E., Stenke, A. and Tummon, F.: Multimodel estimates of atmospheric lifetimes of long-lived ozone-depleting substances: Present and future, J. Geophys. Res. Atmos., 119(5), 2555–2573, doi:10.1002/2013JD021097, 2014.

Deushi, M. and Shibata, K.: Development of a Meteorological Research Institute Chemistry-Climate Model version 2 for the Study of Tropospheric and Stratospheric Chemistry, Pap. Meteorol. Geophys., 62, 1–46, doi:10.2467/mripapers.62.1, 2011.

Duncan, B. N. and Logan, J. A.: Model analysis of the factors regulating the trends and variability of carbon monoxide between 1988 and 1997, Atmos. Chem. Phys., 8(24), 7389–7403, doi:10.5194/acp-8-7389-2008, 2008.

Duncan, B. N., Martin, R. V., Staudt, A. C., Yevich, R. and Logan, J. A.: Interannual and seasonal variability of biomass burning emissions constrained by satellite observations, J. Geophys. Res., 108(D2), 4100, doi:10.1029/2002JD002378, 2003.

Eyring, V. et al.: Overview of IGAC / SPARC Chemistry-Climate Model Initiative (CCMI) Community Simulations in Support of Upcoming Ozone and Climate Assessments, SPARC Newsl., (40)January, 48–66 [online] Available from: http://oceanrep.geomar.de/20227/1/40_SPARCnewsletter_Jan2013_web.pdf, 2013.

Fiore, A. M., Dentener, F. J., Wild, O., Cuvelier, C., Schultz, M. G., Hess, P., Textor, C., Schulz, M., Doherty, R. M., Horowitz, L. W., MacKenzie, I. A., Sanderson, M. G., Shindell, D. T., Stevenson, D. S., Szopa, S., Van Dingenen, R., Zeng, G., Atherton, C., Bergmann, D., Bey, I., Carmichael, G., Collins, W. J., Duncan, B. N., Faluvegi, G., Folberth, G., Gauss, M., Gong, S., Hauglustaine, D., Holloway, T., Isaksen, I. S. A., Jacob, D. J., Jonson, J. E., Kaminski, J. W., Keating, T. J., Lupu, A., Marmer, E., Montanaro, V., Park, R. J., Pitari, G., Pringle, K. J., Pyle, J. A., Schroeder, S., Vivanco, M. G., Wind, P., Wojcik, G., Wu, S. and Zuber, A.: Multimodel estimates of intercontinental source-receptor relationships for ozone pollution, J. Geophys. Res., 114, doi:10.1029/2008jd010816, 2009.

Garcia, R. R., Smith, A. K., Kinnison, D. E., Cámara, Á. de la and Murphy, D. J.: Modification of the Gravity Wave Parameterization in the Whole Atmosphere Community Climate Model: Motivation and Results, J. Atmos. Sci., 74(1), 275–291, doi:10.1175/JAS-D-16-0104.1, 2016.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G. and Yacalis, G.: Could Machine Learning Break the Convection Parameterization Deadlock?, Geophys. Res. Lett., 45(11), 5742–5751, doi:10.1029/2018GL078202, 2018.

Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E. and Hueglin, C.: Random forest meteorological normalisation models for Swiss PM $_{10}$ trend analysis, Atmos. Chem. Phys., 18(9), 6223–6239, doi:10.5194/acp-18-6223-2018, 2018.

Guth, J., Josse, B., Marécal, V., Joly, M. and Hamer, P.: First implementation of secondary inorganic aerosols in the MOCAGE version R2.15.0 chemistry transport model, Geosci. Model Dev., 9(1), 137–160, doi:10.5194/gmd-9-137-2016, 2016.

Hegglin, M. I. and Lamarque, J. F.: The IGAC/SPARC Chemistry-Climate Model Initiative Phase-1 (CCMI-1) model data output, NCAS Br. Atmos. Data Cent. [online] Available from: http://catalogue.ceda.ac.uk/uuid/9cc6b94df0f4469d8066d69b5df879d5 (Accessed 16 July 2019), 2015.

Holmes, C. D.: Methane Feedback on Atmospheric Chemistry: Methods, Models, and Mechanisms, J. Adv. Model. Earth Syst., 10(4), 1087–1099, doi:10.1002/2017MS001196, 2018.

Holmes, C. D., Prather, M. J., Søvde, O. A. and Myhre, G.: Future methane, hydroxyl, and their uncertainties: Key climate and emission parameters for future predictions, Atmos. Chem. Phys., 13, 285–302, doi:10.5194/acp-13-285-2013, 2013.

Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J. and Liu, Y.: Estimating PM 2.5 Concentrations in the Conterminous United States Using the Random Forest Approach, Environ. Sci. Technol., 51(12), 6936–6944, doi:10.1021/acs.est.7b01210, 2017.

Jöckel, P., Brenninkmeijer, C. A. M., Lawrence, M. G., Jeuken, A. B. M. and Velthoven, P. F. J. van: Evaluation of stratosphere–troposphere exchange and the hydroxyl radical distribution in three-dimensional global atmospheric models using observations of cosmogenic ^{14}CO, J. Geophys. Res., 107(D20), 4446, doi:10.1029/2001JD001324, 2002.

Jöckel, P., Tost, H., Pozzer, A., Kunze, M., Kirner, O., Brenninkmeijer, C. A. M., Brinkop, S., Cai, D. S., Dyroff, C., Eckstein, J., Frank, F., Garny, H., Gottschaldt, K.-D., Graf, P., Grewe, V., Kerkweg, A., Kern, B., Matthes, S., Mertens, M., Meul, S., Neumaier, M., Nützel, M., Oberländer-Hayn, S., Ruhnke, R., Runde, T., Sander, R., Scharffe, D. and Zahn, A.: Earth System Chemistry integrated Modelling (ESCiMo) with the Modular Earth Submodel System (MESSy) version 2.51, Geosci. Model Dev., 9(3), 1153–1200, doi:10.5194/gmd-9-1153-2016, 2016.

Josse, B., Simon, P. and Peuch, V.-H.: Radon global simulations with the multiscale chemistry and transport model MOCAGE, Tellus B Chem. Phys. Meteorol., 56(4), 339–356, doi:10.3402/tellusb.v56i4.16448, 2004.

Keller, C. A. and Evans, M. J.: Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10, Geosci. Model Dev., 12(3), 1209–1225, doi:10.5194/gmd-12-1209-2019, 2019.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., Hou, Y. T., Lord, S. J. and Belochitski, A. A.: Accurate and Fast Neural Network Emulations of Model Radiation for the NCEP Coupled Climate Forecast System: Climate Simulations and Seasonal Predictions, Mon. Weather Rev., 138(5), 1822–1842, doi:10.1175/2009MWR3149.1, 2009.

Krol, M., Van Leeuwen, P. J. and Lelieveld, J.: Global OH trend inferred from methylchloroform measurements, J. Geophys. Res., 103711(20), 697–10, 1998.

Lelieveld, J., Gromov, S., Pozzer, A. and Taraborrelli, D.: Global tropospheric hydroxyl distribution, budget and reactivity, Atmos. Chem. Phys., 16(19), 12477–12493, doi:10.5194/acp-16-12477-2016, 2016.

Liang, Q., Chipperfield, M. P., Fleming, E. L., Abraham, N. L., Braesicke, P., Burkholder, J. B., Daniel, J. S., Dhomse, S., Fraser, P. J., Hardiman, S. C., Jackman, C. H., Kinnison, D. E., Krummel, P. B., Montzka, S. A., Morgenstern, O., McCulloch, A., Mühle, J., Newman, P. A., Orkin, V. L., Pitari, G., Prinn, R. G., Rigby, M., Rozanov, E., Stenke, A., Tummon, F., Velders, G. J. M., Visioni, D. and Weiss, R. F.: Deriving Global OH Abundance and Atmospheric Lifetimes for Long-Lived Gases: A Search for CH3CCl3Alternatives, J. Geophys. Res. Atmos., 122(21), 11,914-11,933, doi:10.1002/2017JD026926, 2017.

Liu, Y., Cao, G., Zhao, N., Mulligan, K. and Ye, X.: Improve ground-level PM2.5 concentration mapping using a random forests-based geostatistical approach, Environ. Pollut., 235, 272–282, doi:10.1016/J.ENVPOL.2017.12.070, 2018.

Lovelock, J. E.: Methyl chloroform in the troposphere as an indicator of OH radical abundance, Nature, 267(5606), 32–32, doi:10.1038/267032a0, 1977.

Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N. and Polvani, L. M.: Climate Change from 1850 to 2005 Simulated in CESM1(WACCM), J. Clim., 26(19), 7372–7391, doi:10.1175/JCLI-D-12-00558.1, 2013.

McNorton, J., Chipperfield, M. P., Gloor, M., Wilson, C., Feng, W., Hayman, G. D., Rigby, M., Krumme, P. B., O'Doherty, S., Prinn, R. G., Weiss, R. F., Young, D., Dlugokencky, E. and Montzka, S. A.: Role of OH variability in the stalling of the global atmospheric CH4 growth rate from 1999 to 2006, Atmos. Chem. Phys., doi:10.5194/acp-16-7943-2016, 2016.

Miller, B. R., Huang, J., Weiss, R. F., Prinn, R. G. and Fraser, P. J.: Atmospheric trend and lifetime of chlorodifluoromethane (HCFC-22) and the global tropospheric OH concentration, J. Geophys. Res. Atmos., 103(D11), 13237–13248, doi:10.1029/98JD00771, 1998.

Molod, A., Takacs, L., Suarez, M. and Bacmeister, J.: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2, Geosci. Model Dev., 8(5), 1339–1356, doi:10.5194/gmd-8-1339-2015, 2015.

Montzka, S. A., Spivakovsky, C. M., Butler, J. H., Elkins, J. W., Lock, L. T. and Mondeel, D. J.: New observational constraints for atmospheric hydroxyl on global and hemispheric scales, Science (80-.)., doi:10.1126/science.288.5465.500, 2000.

Morgenstern, O., Hegglin, M., Rozanov, E., O'Connor, F., Luke Abraham, N., Akiyoshi, H., Archibald, A., Bekki, S., Butchart, N., Chipperfield, M., Deushi, M., Dhomse, S., Garcia, R., Hardiman, S., Horowitz, L., Jöckel, P., Josse, B., Kinnison, D., Lin, M., Mancini, E., Manyin, M., Marchand, M., Marécal, V., Michou, M., Oman, L., Pitari, G., Plummer, D., Revell, L., Saint-Martin, D., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tanaka, T., Tilmes, S., Yamashita, Y., Yoshida, K. and Zeng, G.: Review of the global models used within phase 1 of the Chemistry-Climate Model Initiative (CCMI), Geosci. Model Dev., 10(2), 639–671, doi:10.5194/gmd-10-639-2017, 2017.

Naik, V., Voulgarakis, A., Fiore, A. M., Horowitz, L. W., Lamarque, J. F., Lin, M., Prather, M. J., Young, P. J., Bergmann, D., Cameron-Smith, P. J., Cionni, I., Collins, W. J., Dalsoren, S. B., Doherty, R., Eyring, V., Faluvegi, G., Folberth, G. A., Josse, B., Lee, Y. H., MacKenzie, I. A., Nagashima, T., van Noije, T. P. C. C., Plummer, D. A., Righi, M., Rumbold, S. T., Skeie, R., Shindell, D. T., Stevenson, D. S., Strode, S., Sudo, K., Szopa, S., Zeng, G., Dalsøren, S. B., Doherty, R., Eyring, V., Faluvegi, G., Folberth, G. A., Josse, B., Lee, Y. H., MacKenzie, I. A., Nagashima, T., van Noije, T. P. C. C., Plummer, D. A., Righi, M., Rumbold, S. T., Skeie, R., Shindell, D. T., Stevenson, D. S., Strode, S., Sudo, K., Szopa, S. and Zeng, G.: Preindustrial to present-day changes in tropospheric hydroxyl radical and methane lifetime from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., 13(10), 5277–5298, doi:10.5194/acp-13-5277-2013, 2013.

Nicely, J. M., Salawitch, R. J., Canty, T., Anderson, D. C., Arnold, S. R., Chipperfield, M. P., Emmons, L. K., Flemming, J., Huijnen, V., Kinnison, D. E., Lamarque, J. F., Mao, J., Monks, S. A., Steenrod, S. D., Tilmes, S. and Turquety, S.: Quantifying the causes of differences in tropospheric OH within global models, J. Geophys. Res., doi:10.1002/2016JD026239, 2017.

Nicely, J. M., Canty, T. P., Manyin, M., Oman, L. D., Salawitch, R. J., Steenrod, S. D., Strahan, S. E. and Strode, S. A.: Changes in Global Tropospheric OH Expected as a Result of Climate Change Over the Last Several Decades, J. Geophys. Res. Atmos., 123(18), 10,774-10,795, doi:10.1029/2018JD028388, 2018.

Nielsen, J. E., Pawson, S., Molod, A., Auer, B., da Silva, A. M., Douglass, A. R., Duncan, B., Liang, Q., Manyin, M., Oman, L. D., Putman, W., Strahan, S. E. and Wargan, K.: Chemical Mechanisms and Their Applications in the Goddard Earth Observing System (GEOS) Earth System Model, J. Adv. Model. Earth Syst., 9(8), 3019–3044, doi:10.1002/2017MS001011, 2017.

Nisbet, E. G., Dlugokencky, E. J., Manning, M. R., Lowry, D., Fisher, R. E., France, J. L., Michel, S. E., Miller, J. B., White, J. W. C., Vaughn, B., Bousquet, P., Pyle, J. A., Warwick, N. J., Cain, M., Brownlow, R., Zazzeri, G., Lanoisellé, M., Manning, A. C., Gloor, E., Worthy, D. E. J., Brunke, E.-G., Labuschagne, C., Wolff, E. W. and Ganesan, A. L.: Rising atmospheric methane: 2007-2014 growth and isotopic shift, Global Biogeochem. Cycles, 30(9), 1356–1370, doi:10.1002/2016GB005406, 2016.

Nisbet, E. G., Manning, M. R., Dlugokencky, E. J., Fisher, R. E., Lowry, D., Michel, S. E., Myhre, C. L., Platt, S. M., Allen, G., Bousquet, P., Brownlow, R., Cain, M., France, J. L., Hermansen, O., Hossaini, R., Jones, A. E., Levin, I., Manning, A. C., Myhre, G., Pyle, J. A., Vaughn, B., Warwick, N. J. and White, J. W. C.: Very strong atmospheric methane growth in the four years 2014-2017: Implications for the Paris Agreement, Global Biogeochem. Cycles, 2018GB006009, doi:10.1029/2018GB006009, 2019.

Nowack, P., Braesicke, P., Haigh, J., Abraham, N. L., Pyle, J. and Voulgarakis, A.: Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations, Environ. Res. Lett., 13(10), 104016, doi:10.1088/1748-9326/aae2be, 2018.

Orbe, C., Oman, L. D., Strahan, S. E., Waugh, D. W., Pawson, S., Takacs, L. L. and Molod, A. M.: Large-Scale Atmospheric Transport in GEOS Replay Simulations, J. Adv. Model. Earth Syst., 9(7), 2545–2560, doi:10.1002/2017MS001053, 2017.

Prather, M., Ehhalt, D., Dentener, F., Derwent, R., Dlugokencky, E., Holland, E., Isaksen, I., Katima, J., Kirchhoff, V., Matson, P., Midgley, P. and Wang, M.: Atmospheric Chemistry and Greenhouse Gases., 2001.

Prather, M. J. and Holmes, C. D.: Overexplaining or underexplaining methane's role in climate change, Proc. Natl. Acad. Sci., doi:10.1073/pnas.1704884114, 2017.

Prather, M. J., Holmes, C. D. and Hsu, J.: Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry, Geophys. Res. Lett., doi:10.1029/2012GL051440, 2012.

Prinn, R., Cunnold, D., Rasmussen, R., Simmonds, P., Alyea, F., Crawford, A., Fraser, P. and Rosen, R.: Atmospheric Trends in Methylchloroform and the Global Average for the Hydroxyl Radical, Science (80-.)., doi:10.1126/science.238.4829.945, 1987.

Ravishankara, A. R. and Albritton, D. L.: Methyl Chloroform and the Atmosphere, Science (80-.)., 269(5221), 183 LP – 184, doi:10.1126/science.269.5221.183, 1995.

Rigby, M., Montzka, S. A., Prinn, R. G., White, J. W. C., Young, D., O'Doherty, S., Lunt, M. F., Ganesan, A. L., Manning, A. J., Simmonds, P. G., Salameh, P. K., Harth, C. M., Mühle, J., Weiss, R. F., Fraser, P. J., Steele, L. P., Krummel, P. B., McCulloch, A. and Park, S.: Role of atmospheric oxidation in recent methane growth, Proc. Natl. Acad. Sci., doi:10.1073/pnas.1616426114, 2017.

Silva, S. J., Heald, C. L., Ravela, S., Mammarella, I. and Munger, J. W.: A Deep Learning Parameterization for Ozone Dry Deposition Velocities, Geophys. Res. Lett., 46(2), 983–989, doi:10.1029/2018GL081049, 2019.

Singh, H. B.: Atmospheric halocarbons – Evidence in favor of reduced average hydroxyl radical concentration in troposphere, Geophys. Res. Lett., 4(3), 101–104, doi:10.1029/GL004i003p00101, 1977.

Solomon, S., Kinnison, D., Bandoro, J. and Garcia, R.: Simulation of polar ozone depletion: An update, J. Geophys. Res.

Atmos., 120(15), 7958–7974, doi:10.1002/2015JD023365, 2015.

Spivakovsky, C. M., Logan, J. A., Montzka, S. A., Balkanski, Y. J., Foreman-Fowler, M., Jones, D. B. A., Horowitz, L. W., Fusco, A. C., Brenninkmeijer, C. A. M., Prather, M. J., Wofsy, S. C. and McElroy, M. B.: Three-dimensional climatological distribution of tropospheric OH: Update and evaluation, J. Geophys. Res. Atmos., 105(D7), 8931–8980, doi:10.1029/1999JD901006, 2000.

Stauffer, R. M., Thompson, A. M., Oman, L. D. and Strahan, S. E.: The Effects of a 1998 Observing System Change on MERRA-2-Based Ozone Profile Simulations, J. Geophys. Res. Atmos., 0(0), doi:10.1029/2019JD030257, 2019.

Stevenson, D. S., Dentener, F. J., Schultz, M. G., Ellingsen, K., van Noije, T. P. C., Wild, O., Zeng, G., Amann, M., Atherton, C. S., Bell, N., Bergmann, D. J., Bey, I., Butler, T., Cofala, J., Collins, W. J., Derwent, R. G., Doherty, R. M., Drevet, J., Eskes, H. J., Fiore, A. M., Gauss, M., Hauglustaine, D. A., Horowitz, L. W., Isaksen, I. S. A., Krol, M. C., Lamarque, J. F., Lawrence, M. G., Montanaro, V., Muller, J. F., Pitari, G., Prather, M. J., Pyle, J. A., Rast, S., Rodriguez, J. M., Sanderson, M. G., Savage, N. H., Shindell, D. T., Strahan, S. E., Sudo, K. and Szopa, S.: Multimodel ensemble simulations of present-day and near-future tropospheric ozone, J. Geophys. Res., 111(D8), doi:10.1029/2005jd006338, 2006.

Strahan, S. E., Douglass, A. R. and Newman, P. A.: The contributions of chemistry and transport to low arctic ozone in March 2011 derived from Aura MLS observations, J. Geophys. Res. Atmos., 118(3), 1563–1576, doi:10.1002/jgrd.50181, 2013.

Taraborrelli, D., Lawrence, M. G., Crowley, J. N., Dillon, T. J., Gromov, S., Groß, C. B. M., Vereecken, L. and Lelieveld, J.: Hydroxyl radical buffered by isoprene oxidation over tropical forests, Nat. Geosci., 5(3), 190–193, doi:10.1038/ngeo1405, 2012.

Tie, X. and Brasseur, G.: The response of stratospheric ozone to volcanic eruptions: Sensitivity to atmospheric chlorine loading, Geophys. Res. Lett., 22(22), 3035–3038, doi:10.1029/95GL03057, 1995.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Kinnison, D. E., Marsh, D., Garcia, R. R., Smith, A. K., Neely, R. R., Conley, A., Vitt, F., Val Martin, M., Tanimoto, H., Simpson, I., Blake, D. R. and Blake, N.: Representation of the Community Earth System Model (CESM1) CAM4-chem within the Chemistry-Climate Model Initiative (CCMI), Geosci. Model Dev., 9(5), 1853–1890, doi:10.5194/gmd-9-1853-2016, 2016.

Turner, A. J., Frankenberg, C., Wennberg, P. O. and Jacob, D. J.: Ambiguity in the causes for decadal trends in atmospheric methane and hydroxyl, Proc. Natl. Acad. Sci., doi:10.1073/pnas.1616020114, 2017.

Turner, A. J., Frankenberg, C. and Kort, E. A.: Interpreting contemporary trends in atmospheric methane., Proc. Natl. Acad. Sci. U. S. A., 116(8), 2805–2813, doi:10.1073/pnas.1814297116, 2019.

Verstraeten, W. W., Neu, J. L., Williams, J. E., Bowman, K. W., Worden, J. R. and Boersma, K. F.: Rapid increases in tropospheric ozone production and export from China, Nat. Geosci., 8, 690 [online] Available from: https://doi.org/10.1038/ngeo2493, 2015.

Voulgarakis, A., Naik, V., Lamarque, J. F., Shindell, D. T., Young, P. J., Prather, M. J., Wild, O., Field, R. D., Bergmann, D., Cameron-Smith, P., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R. M., Eyring, V., Faluvegi, G., Folberth, G. A., Horowitz, L. W., Josse, B., MacKenzie, I. A., Nagashima, T., Plummer, D. A., Righi, M., Rumbold, S. T., Stevenson, D. S., Strode, S. A., Sudo, K., Szopa, S., Zeng, G., Dalsoren, S. B., Doherty, R. M., Eyring, V., Faluvegi, G., Folberth, G. A., Horowitz, L. W., Josse, B., MacKenzie, I. A., Nagashima, T., Plummer, D. A., Righi, M., Rumbold, S. T., Stevenson, D. S., Strode, S. A., Sudo, K., Szopa, S. and Zeng, G.: Analysis of present day and future OH and methane lifetime in the ACCMIP simulations, Atmos. Chem. Phys., 13(5), 2563–2587, doi:10.5194/acp-13-2563-2013, 2013.

Wargan, K., Orbe, C., Pawson, S., Ziemke, J. R., Oman, L. D., Olsen, M. A., Coy, L. and Emma Knowland, K.: Recent Decline in Extratropical Lower Stratospheric Ozone Attributed to Circulation Changes, Geophys. Res. Lett., 45(10), 5166–5176, doi:10.1029/2018GL077406, 2018.

Weinstock, B. and Niki, H.: Carbon Monoxide: Residence Time in the Atmosphere, Science (80-.)., 166(3902), 224–225, doi:10.1126/science.166.3902.224, 1969.

Wild, O.: Modelling the global tropospheric ozone budget: exploring the variability in current models, Atmos. Chem. Phys., 7(10), 2643–2660, doi:10.5194/acp-7-2643-2007, 2007.

YUKIMOTO, S., ADACHI, Y., HOSAKA, M., SAKAMI, T., YOSHIMURA, H., HIRABARA, M., TANAKA, T. Y., SHINDO, E., TSUJINO, H., DEUSHI, M., MIZUTA, R., YABU, S., OBATA, A., NAKANO, H., KOSHIRO, T., OSE, T. and KITOH, A.: A New Global Climate Model of the Meteorological Research Institute: MRI-CGCM3 –Model Description and Basic Performance–, J. Meteorol. Soc. Japan, 90A(0), 23–64, doi:10.2151/jmsj.2012-A02, 2012.

Zeng, G., Wood, S. W., Morgenstern, O., Jones, N. B., Robinson, J. and Smale, D.: Trends and variations in CO, C2H6, and HCN in the Southern Hemisphere point to the declining anthropogenic emissions of CO and C2H6, Atmos. Chem. Phys., 12(16), 7543–7555, doi:10.5194/acp-12-7543-2012, 2012.

Zhao, Y., Saunois, M., Bousquet, P., Lin, X., Berchet, A., Hegglin, M. I., Canadell, J. G., Jackson, R. B., Hauglustaine, D.

A., Szopa, S., Stavert, A. R., Abraham, N. L., Archibald, A. T., Bekki, S., Deushi, M., Jöckel, P., Josse, B., Kinnison, D., Kirner, O., Marécal, V., O&apos;Connor, F. M., Plummer, D. A., Revell, L. E., Rozanov, E., Stenke, A., Strode, S., Tilmes, S., Dlugokencky, E. J. and Zheng, B.: Inter-model comparison of global hydroxyl radical (OH) distributions and their impact on atmospheric methane over the 2000–2016 period, Atmos. Chem. Phys., 19(21), 13701–13723, doi:10.5194/acp-19-13701-2019, 2019.

Table 1. Accounting of CH_4 lifetime differences between GMI and OsloCTM simulations for January, 2000.

		GMI	OsloCTM
$\tau_{CH4,ORIG}$[a] (year)		9.24	7.18
$\Delta\tau_{CH4}$ due to[b]:	O_3	–0.91	+0.79
	JO^1D	–0.59	+0.60
	HCHO	–0.64	+0.51
	NO_x	–0.45	+0.33
	JNO_2	–0.34	+0.15
	Isoprene	–0.03	+0.28
	CO	+0.19	–0.07
	H_2O	+0.10	–0.13
	CH_4^{NORM}	+0.11	–0.06
	NO/NO_x	+0.07	–0.05
	O_3 COL	–0.02	–0.06
	T	–0.02	+0.00
$\Delta\tau_{CH4,TOT}$[c]		–2.52	+2.30
$\tau_{CH4,ORIG} + \Delta\tau_{CH4,TOT}$		6.71	9.48
Mech.[d]		+0.47	–0.24

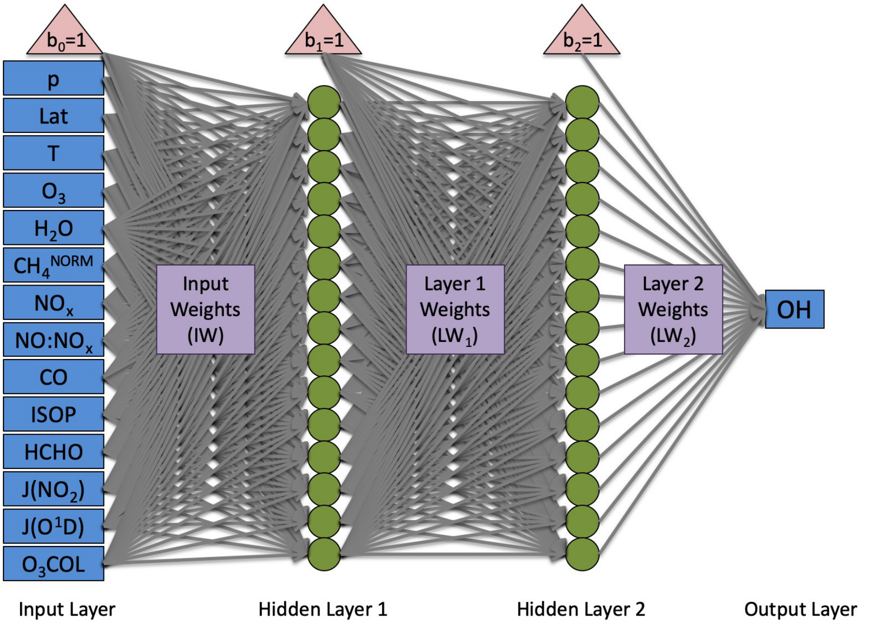
[a] $\tau_{CH4,ORIG}$ represents value of τ_{CH4} evaluated directly from the model.

[b] $\Delta\tau_{CH4}$ calculated from output of NN when noted variable is substituted with values from the other model.

[c] Sum of all $\Delta\tau_{CH4}$ values calculated for each input substitution.

[d] Remainder of original τ_{CH4} difference not accounted for by NN substitutions; calculated as $\tau_{CH4,ORIG}$(model A) – [$\tau_{CH4,ORIG}$(model B) + $\Delta\tau_{CH4,TOT}$(model B)].

Formatted: Superscript



$$OH = b_2 \times LW_2 + \Sigma[LW_2 \times \tanh(b_1 \times LW_1 + \Sigma[LW_1 \times \tanh(b_0 \times IW + \Sigma[IW \times inputs])])]$$

Figure 1. Architecture for neural networks generated in this study. Blue boxes designate inputs (left) and output (right), red triangles indicate bias terms, green circles indicate nodes at which activation functions are performed, and grey arrows represent the weighting terms, which are optimized through the training process. For full details of the neural network setup and training, we refer readers to Nicely et al. [2017]. Although 15 nodes are shown here in each hidden layer, 30 were actually used for all NNs in this study.

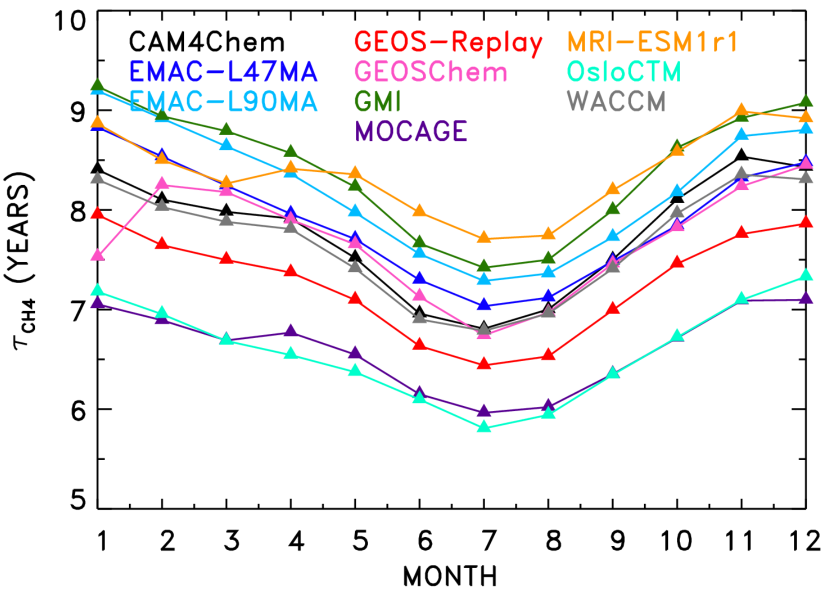


Figure 2. Seasonal variation in CH_4 lifetime for year 2000 for the CCMI specified dynamics (REF-C1SD) and chemical transport model simulations.

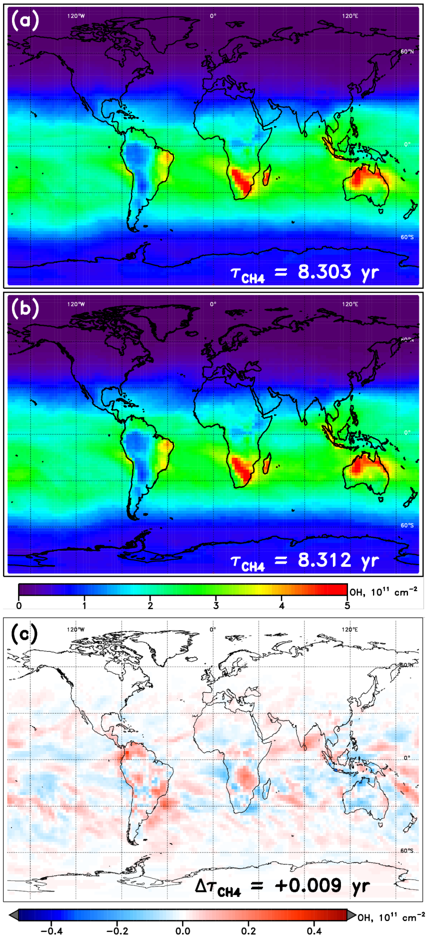


Figure 3. Tropospheric OH columns for the WACCM model REF-C1SD simulation, January 2000. (a) Columns calculated directly from the WACCM output; (b) columns calculated from the output from the WACCM January NN run with inputs from the native model; (c) difference in column values, (NN – model). Methane lifetime values calculated from 3-D OH fields from WACCM and from the WACCM NN are inscribed in panels (a) and (b), respectively. The methane lifetime difference, (NN – model), is noted in panel (c).

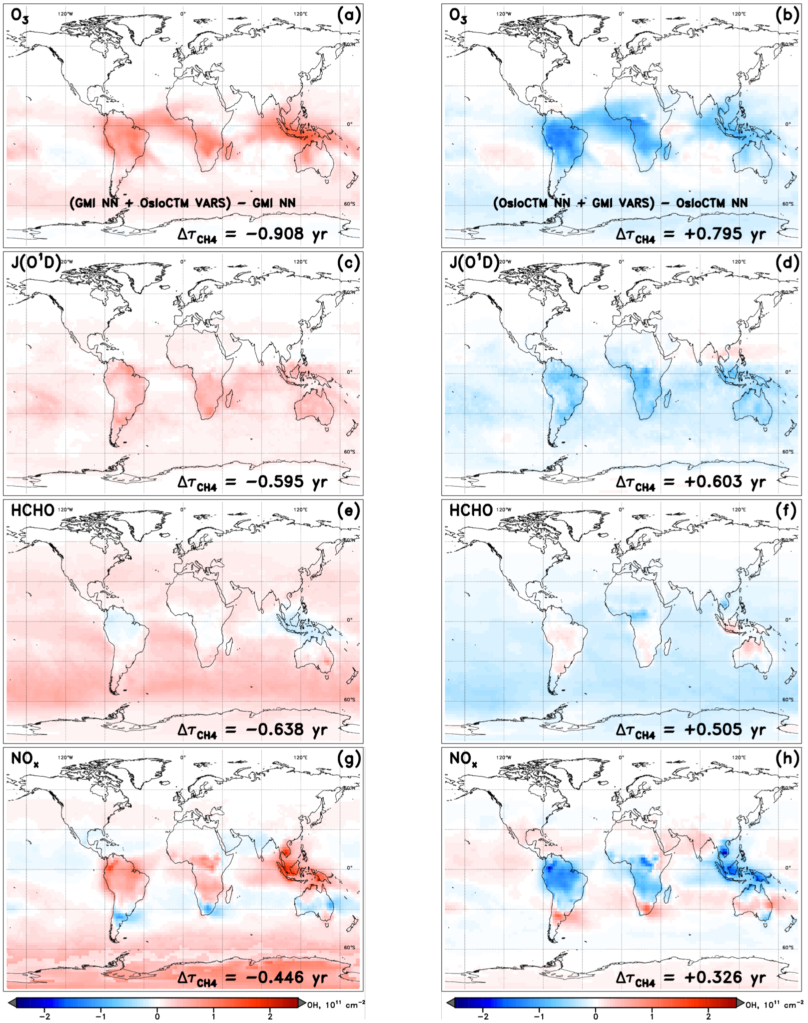


Figure 4. Changes in tropospheric OH column resulting from swap of indicated variable from another model into the NN of the native model for the specified dynamics simulation of January, 2000. Swaps of the inputs O_3 (a, b), $J(O_3 \rightarrow O^1D)$ (c, d), HCHO (e, f), and NO_x (g, h) are shown for the GMI (left) and OsloCTM (right) NNs.

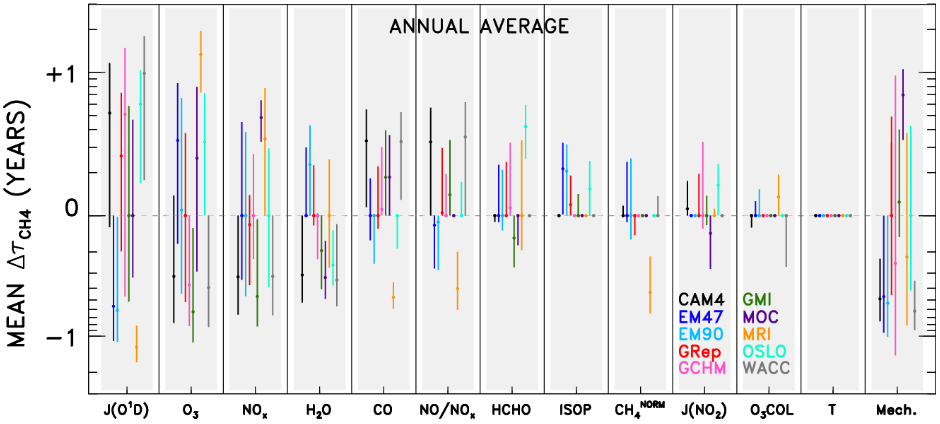


Figure 5. Averaged changes in CH_4 lifetime accrued for a specified model (color), across all swaps of the indicated variable (x-axis) from all other models. Results are shown annually averaged for year 2000 of the specified dynamics REF-C1SD CCMI and chemical transport model simulations. Circle indicates the mean change in CH_4 lifetime; bars represent the 1σ standard deviation from all model pairings. Variables along the x-axis are ranked by averaged magnitude of the $\Delta\tau_{CH4}$ values (i.e., inputs located farther left are responsible for larger differences in CH_4 lifetime), except for the "Mech.+Nonlin." term, which is shown last to indicate its role as a remainder term. Model name abbreviations are "CAM4" for CAM4-Chem, "EM47" for EMAC-L47MA, "EM90" for EMAC-L90MA, "GRep" for GEOS Replay, "GCHM" for GEOS-Chem, "GMI" for GMI, "MOC" for MOCAGE, "MRI" for MRI-ESM1r1, "OSLO" for OsloCTM, and "WACC" for WACCM.

1035

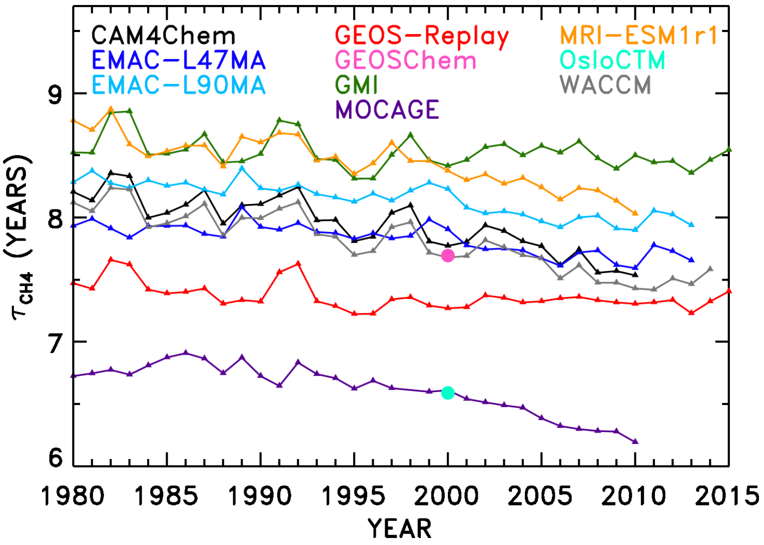


Figure 6. Time series of CH_4 lifetime from REF-C1SD models. Only one year of output was available for two models (OsloCTM and GEOS-Chem), so their results are shown only as a single data point at year 2000.

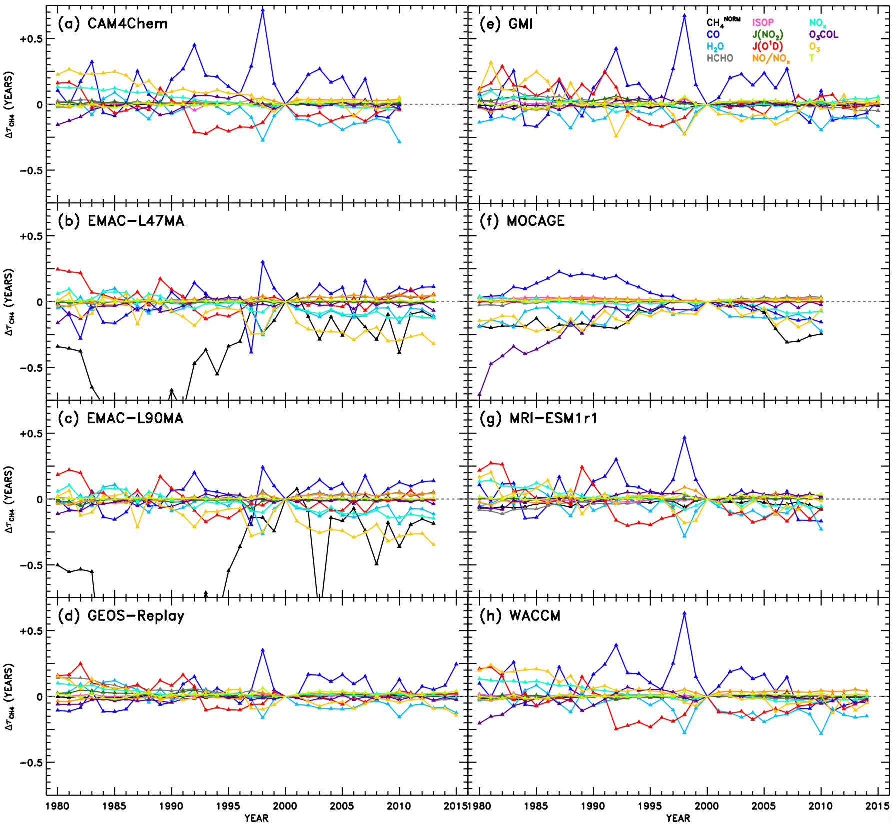


Figure 7. **Attributions of changes in CH_4 lifetime relative to year 2000 of the REF-C1SD simulations. Within the NN of a given model, use of individual inputs (indicated by color) from years other than 2000 result in a change and OH and subsequent CH_4 lifetime, shown here. The variations attributable to CH_4 are labeled "CH_4^{NORM} to designate the use of normalized CH_4 fields as inputs to the NNs, as described in Sections 3.1 and 4.3. As a result, OH changes due to CH_4^{NORM} represent impacts of changes in how CH_4 is distributed within the troposphere, rather than how CH_4 concentrations are changing over time.**

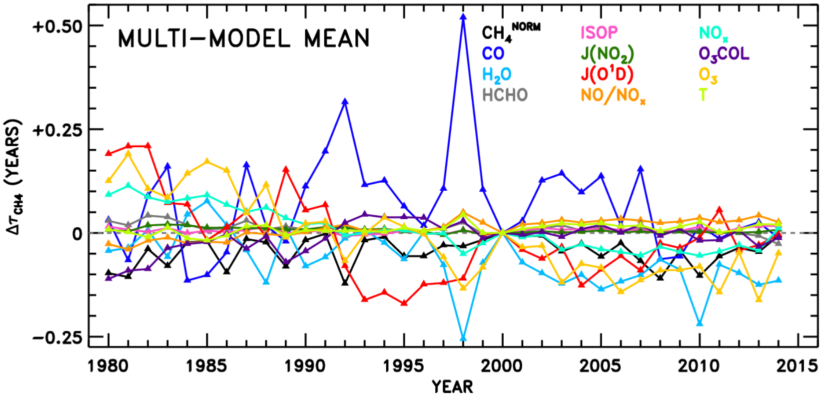


Figure 8. Same as Figure 7, but the average across all eight models, except filtered to remove NN results for individual months and years during which NN performance is poor, as detailed in the text.

Deleted: with the CH_4 line from EMAC-L47MA and EMAC-L90MA and O_3 Column from MOCAGE removed…

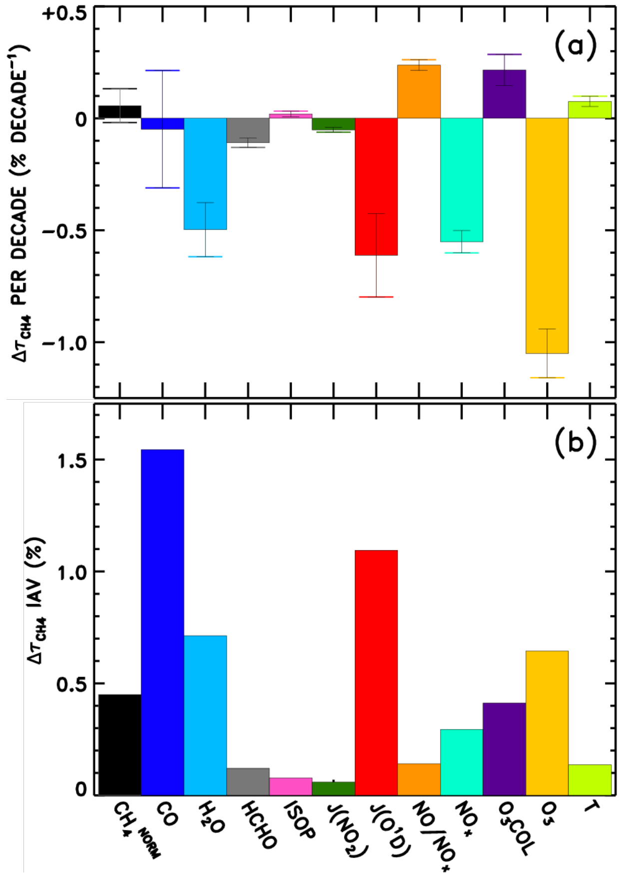


Figure 9. Multi-model mean linear trend (a) and interannual variability (b) in τ_{CH4} attributed to each variable examined through the NN method.

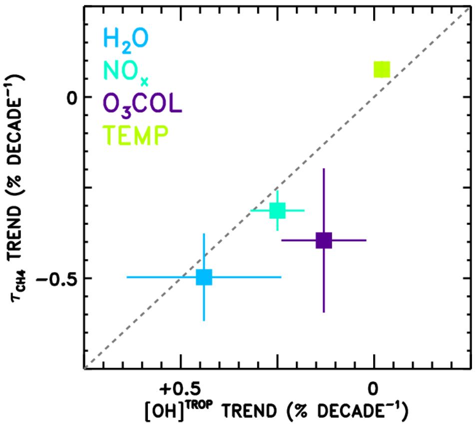


Figure 10. Comparison of the attributed trends in τ_{CH4} found in this work according to the REF-C1SD simulations performed for CCMI (y-axis) to the attributed trends in tropospheric mean OH ("$[OH]^{TROP}$") found based on observations in Nicely et al. [2018]. The grey dashed line indicates the –1:1 line, as values should be anti-correlated. The τ_{CH4} trend numbers from this work for NO_x combine the NO_x total abundance and partitioning (NO/NO_x) values from Figure 9, and for O_3 Column combine the $J(O^1D)$ and O_3 Column values, as both effects are encompassed in the determination of $[OH]^{TROP}$.

Supporting Information for:

A Machine Learning Examination of Hydroxyl Radical Differences Among Model Simulations for CCMI-1

Julie M. Nicely[1,2], Bryan N. Duncan[2], Thomas F. Hanisco[2], Glenn M. Wolfe[2,3] Ross J. Salawitch[1,4,5], Makoto Deushi[6], Amund S. Haslerud[7], Patrick Jöckel[8], Béatrice Josse[9], Douglas E. Kinnison[10], Andrew Klekociuk[11,12], Michael E. Manyin[2,13], Virginie Marécal[9], Olaf Morgenstern[14], Lee T. Murray[15], Gunnar Myhre[7], Luke D. Oman[2], Giovanni Pitari[16], Andrea Pozzer[17], Ilaria Quaglia[16], Laura E. Revell[18], Eugene Rozanov[19,20], Andrea Stenke[19], Kane Stone[21,22], Susan Strahan[2,23], Simone Tilmes[10], Holger Tost[24], Daniel M. Westervelt[25,26], and Guang Zeng[14]

[1]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA.
[2]NASA Goddard Space Flight Center, Greenbelt, MD, USA.
[3]Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore, MD, USA.
[4]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, USA.
[5]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD, USA.
[6]Meteorological Research Institute (MRI), Tsukuba, Japan.
[7]Center for International Climate and Environmental Research-Oslo (CICERO), Oslo, Norway.
[8]Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, Germany.
[9]CNRM UMR 3589, Météo-France/CNRS, Toulouse, France.
[10]National Center for Atmospheric Research, Boulder, CO, USA.
[11]Antarctica and the Global System Program, Australian Antarctic Division, Kingston, Australia.
[12]Antarctic Climate and Ecosystems Cooperative Research Centre, Hobart, Australia.
[13]Science Systems and Applications, Inc., Lanham, MD, USA.
[14]National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand.
[15]Department of Earth and Environmental Sciences, University of Rochester, Rochester, NY, USA.
[16]Department of Physical and Chemical Sciences, Universitá dell'Aquila, L'Aquila, Italy.
[17]Max-Planck-Institute for Chemistry, Air Chemistry Department, Mainz, Germany.
[18]School of Physical and Chemical Sciences, University of Canterbury, Christchurch, New Zealand.
[19]Institute for Atmospheric and Climate Science, ETH Zürich (ETHZ), Zürich, Switzerland.
[20]Physikalisch-Meteorologisches Observatorium Davos – World Radiation Center (PMOD/WRC), Davos, Switzerland.
[21]School of Earth Sciences, University of Melbourne, Melbourne, Australia.
[22]Massachusetts Institute of Technology, Cambridge, MA, USA.
[23]Universities Space Research Association, Columbia, MD, USA.
[24]Institute for Atmospheric Physics, Johannes Gutenberg University, Mainz, Germany.
[25]Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA.
[26]NASA Goddard Institute for Space Studies, New York, NY, USA.

Correspondence to: Julie M. Nicely (julie.m.nicely@nasa.gov)

Contents:

Text S1
Tables S1-S3
Figures S1-S25

Deleted: 11

Text S1

Here we present and discuss our analysis of the REF-C1 historical free-running simulations from CCMI. These simulations differ from those presented in the main body of the text in that the models do not constrain their meteorological fields in any way to historical meteorology. Winds, temperature, pressure, and water vapor are internally calculated by the Chemistry Climate Models (CCMs), so it is unlikely that meteorological features, such as ENSO and drought-induced biomass burning, align with reality.

The CCMs that provided REF-C1 simulations, including all output necessary to perform the same NN training and inter-model comparison described in the main text (Sections 3.1 and 3.2), are: ACCESS-CCM, CAM4-Chem, EMAC-L47MA, EMAC-L90MA, GEOSCCM, MOCAGE, MRI-ESM1r1, NIWA-UKCA, SOCOL3, ULAQ-CCM, and WACCM. Details of the REF-C1 simulation, performed for 1960-2010, are found in Hegglin & Lamarque (2015) and Morgenstern et al. (2017). One model, the Coupled Model (CM3) developed at the Geophysical Fluid Dynamics Laboratory (GFDL) (Donner et al., 2011) is added to the free-running analysis. The simulation of the CM3 model used here is a 400-year time-slice run, with perpetual emissions representative of year 2000 (Westervelt et al., 2018). Further details of the model setup are available in Westervelt et al. (2017). By including CM3 with the group of REF-C1 CCMI models, we analyse a total of 12 free-running models.

The inter-model comparison conducted for the REF-C1 model simulations was performed following the same protocol as described in Section 3.2 of the main text. The values of $\boldsymbol{\tau}_{\mathbf{CH_4}}$ calculated for each month of year 2000 are shown in **Figure S26**, while the annual average changes in $\boldsymbol{\tau}_{\mathbf{CH_4}}$ ($\boldsymbol{\Delta\tau}_{\mathbf{CH4}}$) by model, for NN swaps of the indicated species, are shown in **Figure S27**. Overall, values of $\boldsymbol{\Delta\tau}_{\mathbf{CH4}}$ are larger than the same values calculated for the REF-C1SD specified dynamics simulations examined in the main text, and chemical mechanism differences appear to play a larger role. For example, the variables responsible for the largest OH differences are O_3 in the free-running simulations and JO^1D in the specified dynamics simulations (Fig. 5). The mean absolute value of the annual average $\boldsymbol{\Delta\tau}_{\mathbf{CH4}}$ due to O_3 in the free-running models is 0.60 ± 0.69 years, while the same aggregation of $\boldsymbol{\Delta\tau}_{\mathbf{CH4}}$ values due to JO^1D in the specified dynamics models is 0.54 ± 0.57 years. The second-most important variables, NO_x in the free-running simulations and O_3 in the specified dynamics simulations, yield $\boldsymbol{\Delta\tau}_{\mathbf{CH4}}$ values of, on average, 0.48 ± 1.11 years and 0.42 ± 0.49 years, respectively. The remainder $\boldsymbol{\Delta\tau}_{\mathbf{CH4}}$ attributed to chemical mechanism differences between models averages to 0.69 ± 1.14 years in the free-running simulations as opposed to 0.36 ± 0.46 years in the specified dynamics simulations.

Deleted: 9

Deleted: 10

The larger values of $\boldsymbol{\Delta\tau}_{\mathbf{CH4}}$ in the free-running models may convey that meteorological differences are imparting an impact on OH through mechanisms that are not sufficiently represented in the input variables chosen for the NN analysis. It is possible that other chemical species not included here that are substantially altered by meteorology or transport and in turn alter OH concentrations would manifest as larger values of $\boldsymbol{\Delta\tau}_{\mathbf{CH4}}$, particularly in the Mech. term. On the other hand, if those missing species are correlated with one of the species or variables used as an input to the NN, the $\boldsymbol{\Delta\tau}_{\mathbf{CH4}}$ attributed to that input may also be inflated. As a result, we caution that model variations in meteorological conditions, expected as a result of their free-running setup in the REF-C1 simulation, could generate artifacts that are less likely to arise in the REF-C1SD simulation comparison, in which temperatures, transport, cloud cover, and water vapor should be reasonably similar.

As with the inter-model comparison of the specified dynamics simulations, results of the free-running model analysis exhibit a multitude of interesting features. While we cannot explore each one with the amount of attention it is due, we would like to discuss one example that highlights the utility of the NN method. In Fig. S27, the $\boldsymbol{\Delta\tau_{CH4}}$ attributed to JNO_2 shows curious behavior for the SOCOL3 model. The absence of spread about the mean value of $\boldsymbol{\Delta\tau_{CH4}}$ is highly unusual, except for instances where a model shows no or very little response of OH to a NN input. The relatively large value of $\boldsymbol{\Delta\tau_{CH4}}$ for SOCOL3 (+0.69±0.09 years) paired with the small variation in this quantity across all the model pairings most likely indicates an issue in the model. **Figure S28** shows the JNO_2 fields, taken directly from each CCM for January, 2000, at 850 hPa. There is much diversity in this quantity across all the models, but the SOCOL3 model exhibits markedly high values, within the tropics especially. Revell et al. (2018) also identify this issue and suggest that the treatment of solar backscatter from clouds may be responsible for biases in the photolysis look-up table calculations. Additionally, a geometric spatial pattern is evident between the latitudes 0° and 30°S, which is unlikely to result from any physical process in the true atmosphere. This may indicate a problem in the way time averaging is conducted to achieve the monthly mean fields reported, a dependence within the photolysis code on a non-continuous time variable (since the pattern repeats regularly every 30° of longitude), or a similar issue. To reduce the likelihood of a bias due to differences in the way that monthly means are calculated, it may be useful for future inter-model comparison efforts to clearly define a desired method of averaging (e.g., composing daily averages from hourly output then averaging the daily means as opposed to averaging a month's worth of 6-hourly instantaneous output). It is of course possible to identify this variety of idiosyncrasy by careful inspection of each model field that is output from a model, but that is a time- and labor-intensive task. Instead, the NN method is capable of pointing a user directly to the offending fields, at least for the variables that are of sufficient relevance to OH chemistry that we have included them here as inputs. In the case that a user wants simply to detect outlier model fields as in this case, it is entirely feasible that the NN method could be adapted for that purpose.

Deleted: 10

Deleted: 11

References

Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., Golaz, J. C., Ginoux, P., Lin, S. J., Schwarzkopf, M. D., Austin, J., Alaka, G., Cooke, W. F., Delworth, T. L., Freidenreich, S. M., Gordon, C. T., Griffies, S. M., Held, I. M., Hurlin, W. J., Klein, S. A., Knutson, T. R., Langenhorst, A. R., Lee, H. C., Lin, Y., Magi, B. I., Malyshev, S. L., Milly, P. C. D., Naik, V., Nath, M. J., Pincus, R., Ploshay, J. J., Ramaswamy, V., Seman, C. J., Shevliakova, E., Sirutis, J. J., Stern, W. F., Stouffer, R. J., Wilson, R. J., Winton, M., Wittenberg, A. T. and Zeng, F.: The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3, J. Clim., 24(13), 3484–3519, doi:10.1175/2011JCLI3955.1, 2011.

Hegglin, M. I. and Lamarque, J. F.: The IGAC/SPARC Chemistry-Climate Model Initiative Phase-1 (CCMI-1) model data output, NCAS Br. Atmos. Data Cent. [online] Available from: http://catalogue.ceda.ac.uk/uuid/9cc6b94df0f4469d8066d69b5df879d5 (Accessed 16 July 2019), 2015.

Morgenstern, O., Hegglin, M., Rozanov, E., O'Connor, F., Luke Abraham, N., Akiyoshi, H., Archibald, A., Bekki, S., Butchart, N., Chipperfield, M., Deushi, M., Dhomse, S., Garcia, R., Hardiman, S., Horowitz, L., Jöckel, P., Josse, B., Kinnison, D., Lin, M., Mancini, E., Manyin, M., Marchand, M., Marécal, V., Michou, M., Oman, L., Pitari, G., Plummer, D., Revell, L., Saint-Martin, D., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tanaka, T., Tilmes, S., Yamashita, Y., Yoshida, K. and Zeng, G.: Review of the global models used within phase 1 of the Chemistry-Climate Model Initiative (CCMI), Geosci. Model Dev., 10(2), 639–671, doi:10.5194/gmd-10-639-2017, 2017.

Revell, L. E., Stenke, A., Tummon, F., Feinberg, A., Rozanov, E., Peter, T., Abraham, N. L., Akiyoshi, H., Archibald, A. T., Butchart, N., Deushi, M., Jöckel, P., Kinnison, D., Michou, M., Morgenstern, O., O&apos;Connor, F. M., Oman, L. D., Pitari, G., Plummer, D. A., Schofield, R., Stone, K., Tilmes, S., Visioni, D., Yamashita, Y. and Zeng, G.: Tropospheric ozone in CCMI models and Gaussian process emulation to understand biases in the SOCOLv3 chemistry–climate model, Atmos. Chem. Phys., 18(21), 16155–16172, doi:10.5194/acp-18-16155-2018, 2018.

Westervelt, D. M., Conley, A. J., Fiore, A. M., Lamarque, J.-F., Shindell, D., Previdi, M., Faluvegi, G., Correa, G. and Horowitz, L. W.: Multimodel precipitation responses to removal of U.S. sulfur dioxide emissions, J. Geophys. Res. Atmos., 122(9), 5024–5038, doi:10.1002/2017JD026756, 2017.

Westervelt, D. M., Conley, A. J., Fiore, A. M., Lamarque, J.-F., Shindell, D. T., Previdi, M., Mascioli, N. R., Faluvegi, G., Correa, G. and Horowitz, L. W.: Connecting regional aerosol emissions reductions to local and remote precipitation responses, Atmos. Chem. Phys., 18(16), 12461–12475, doi:10.5194/acp-18-12461-2018, 2018.

Table S1. The sample size, i.e., number of tropospheric model gridboxes, used for the training of each model's NN, for each month

Model	**January**	**April**	**July**	**October**
CAM4Chem	424,381	422,602	430,701	427,699
EMAC-L47MA	179,048	178,435	184,163	182,065
EMAC-L90MA	167,025	166,307	173,231	170,621
GEOS-Replay	396,694	395,066	401,571	398,479
GEOS-Chem	101,305	100,587	102,441	101,758
GMI	398,655	397,802	401,866	400,860
MOCAGE	427,194	426,621	430,437	432,303
MRI-ESM1r1	270,651	267,521	284,988	273,127
OsloCTM	378,332	376,794	382,529	379,338
WACCM	424,298	422,580	430,601	427,766

Formatted: Left

Formatted Table

Formatted: Font: Not Bold

Table S2. Constrained meteorology (REF-C1SD) CCMI model neural network training statistics, including mean squared error with regularization (MSE_REG), mean squared error (MSE), sum of squared errors (SSE) and Pearson correlation coefficient (R).

Model	**Month**	**MSE_REG**	**MSE**	**SSE**	**R**
CAM4Chem	1	0.37923	4.1261×10^{-4}	175.11	0.99985
	4	0.29407	2.1851×10^{-4}	92.341	0.99987
	7	0.10253	3.2249×10^{-4}	138.90	0.99991
	10	0.19263	2.5806×10^{-4}	110.37	0.99980
EMAC-L47MA	1	0.18339	6.7995×10^{-4}	121.74	0.99976
	4	0.18699	2.0219×10^{-4}	36.078	0.99986
	7	0.087789	3.4581×10^{-4}	63.685	0.99991
	10	0.17907	1.9427×10^{-4}	35.369	0.99983
EMAC-L90MA	1	0.35482	5.2862×10^{-4}	88.293	0.99981
	4	0.099595	2.0500×10^{-4}	34.093	0.99985
	7	0.080911	4.1468×10^{-4}	71.836	0.99989
	10	0.20788	1.7210×10^{-4}	29.364	0.99985
GEOS-Replay	1	0.20707	2.6069×10^{-4}	103.41	0.99990
	4	0.18929	1.2464×10^{-4}	49.240	0.99991
	7	0.080128	2.8015×10^{-4}	112.50	0.99992
	10	0.22672	1.5059×10^{-4}	60.009	0.99988
GEOSChem	1	0.22284	1.5205×10^{-4}	15.404	0.99993
	4	0.061255	1.1911×10^{-4}	11.981	0.99994
	7	0.14682	1.3059×10^{-4}	13.378	0.99997
	10	0.14627	1.0033×10^{-4}	10.210	0.99993
GMI	1	0.074667	4.8862×10^{-4}	194.79	0.99980
	4	0.094182	1.8318×10^{-4}	72.869	0.99987
	7	0.11295	3.1920×10^{-4}	128.27	0.99990
	10	0.14209	2.0734×10^{-4}	83.115	0.99984
MOCAGE	1	0.71640	9.0308×10^{-4}	385.79	0.99969
	4	0.48984	5.0834×10^{-4}	216.87	0.99961
	7	0.34163	8.6047×10^{-4}	370.38	0.99976
	10	0.31921	6.3376×10^{-4}	273.98	0.99944
MRI-ESM1r1	1	0.23535	5.9525×10^{-4}	161.10	0.99972
	4	0.10819	4.0592×10^{-4}	108.59	0.99975
	7	0.099627	6.2950×10^{-4}	179.40	0.99986
	10	0.14278	3.6378×10^{-4}	99.36	0.99971
OsloCTM	1	0.099237	7.7313×10^{-4}	292.50	0.99972
	4	0.19119	2.1283×10^{-4}	80.193	0.99986
	7	0.17034	4.0701×10^{-4}	155.69	0.99988
	10	0.16749	1.9759×10^{-4}	74.955	0.99984
WACCM	1	0.075339	5.1098×10^{-4}	216.81	0.99980
	4	0.23018	1.9571×10^{-4}	82.702	0.99986
	7	0.11848	2.4019×10^{-4}	103.43	0.99991
	10	0.087687	3.0286×10^{-4}	129.55	0.99974

Table S3. Free-running (REF-C1) CCMI model neural network training statistics, as in Table S2.

Model	**Month**	**MSE_REG**	**MSE**	**SSE**	**R**
ACCESS-CCM	1	0.11715	9.6330×10^{-4}	163.80	0.99968
	4	0.16688	2.3802×10^{-4}	40.386	0.99987
	7	0.19878	3.3644×10^{-4}	58.507	0.99991
	10	0.089005	3.6215×10^{-4}	62.611	0.99974
CAM4Chem	1	0.055129	6.6306×10^{-4}	127.79	0.99974
	4	0.11032	2.8507×10^{-4}	54.614	0.99980
	7	0.13067	3.8054×10^{-4}	75.732	0.99988
	10	0.12742	2.5473×10^{-4}	49.706	0.99978
CM3	1	0.11112	4.6630×10^{-4}	125.73	0.99980
	4	0.16528	1.8858×10^{-4}	50.816	0.99989
	7	0.026817	8.8304×10^{-4}	239.79	0.99980
	10	0.082181	3.2942×10^{-4}	91.146	0.99971
EMAC-L47MA	1	0.072276	9.2591×10^{-4}	171.04	0.99969
	4	0.086863	2.4612×10^{-4}	45.659	0.99983
	7	0.10271	4.7184×10^{-4}	90.056	0.99988
	10	0.11785	2.0844×10^{-4}	39.311	0.99982
EMAC-L90MA	1	0.16911	7.2262×10^{-4}	126.75	0.99976
	4	0.076460	2.3894×10^{-4}	42.177	0.99982
	7	0.081041	5.0940×10^{-4}	92.539	0.99988
	10	0.12357	2.2109×10^{-4}	39.097	0.99982
GEOSCCM	1	0.18199	3.9622×10^{-4}	158.05	0.99984
	4	0.064835	2.3637×10^{-4}	94.576	0.99984
	7	0.18798	3.2447×10^{-4}	130.95	0.99991
	10	0.11525	2.5013×10^{-4}	100.29	0.99982
MOCAGE	1	0.34317	1.3971×10^{-3}	617.16	0.99957
	4	0.70165	4.7860×10^{-4}	210.00	0.99963
	7	0.54001	9.6137×10^{-4}	431.00	0.99976
	10	0.31103	5.6591×10^{-4}	254.71	0.99938
MRI-ESM1r1	1	0.25895	6.1599×10^{-4}	166.11	0.99970
	4	0.20258	3.8259×10^{-4}	101.85	0.99977
	7	0.22064	4.3767×10^{-4}	121.15	0.99987
	10	0.24607	3.4118×10^{-4}	92.663	0.99971
NIWA-UKCA	1	0.065453	1.1373×10^{-3}	193.66	0.99959
	4	0.17948	2.5409×10^{-4}	43.007	0.99986
	7	0.13032	3.8962×10^{-4}	68.185	0.99989
	10	0.18349	2.9323×10^{-4}	50.539	0.99981
SOCOL3	1	0.10894	7.6744×10^{-4}	85.386	0.99940
	4	0.15187	1.6699×10^{-4}	18.597	0.99978
	7	0.16907	3.5237×10^{-4}	40.847	0.99985
	10	0.14254	2.4928×10^{-4}	28.546	0.99953
ULAQ-CCM	1	0.058951	2.7677×10^{-4}	12.950	0.99989
	4	0.047113	2.0840×10^{-4}	9.6693	0.99982
	7	0.057370	2.9026×10^{-4}	14.398	0.99990

	10	0.065023	1.3451×10^{-4}	6.5473	0.99988
WACCM	1	0.050710	5.7059×10^{-4}	110.51	0.99975
	4	0.11753	1.9040×10^{-4}	36.453	0.99984
	7	0.053290	3.2026×10^{-4}	63.519	0.99988
	10	0.13621	2.3013×10^{-4}	44.987	0.99977

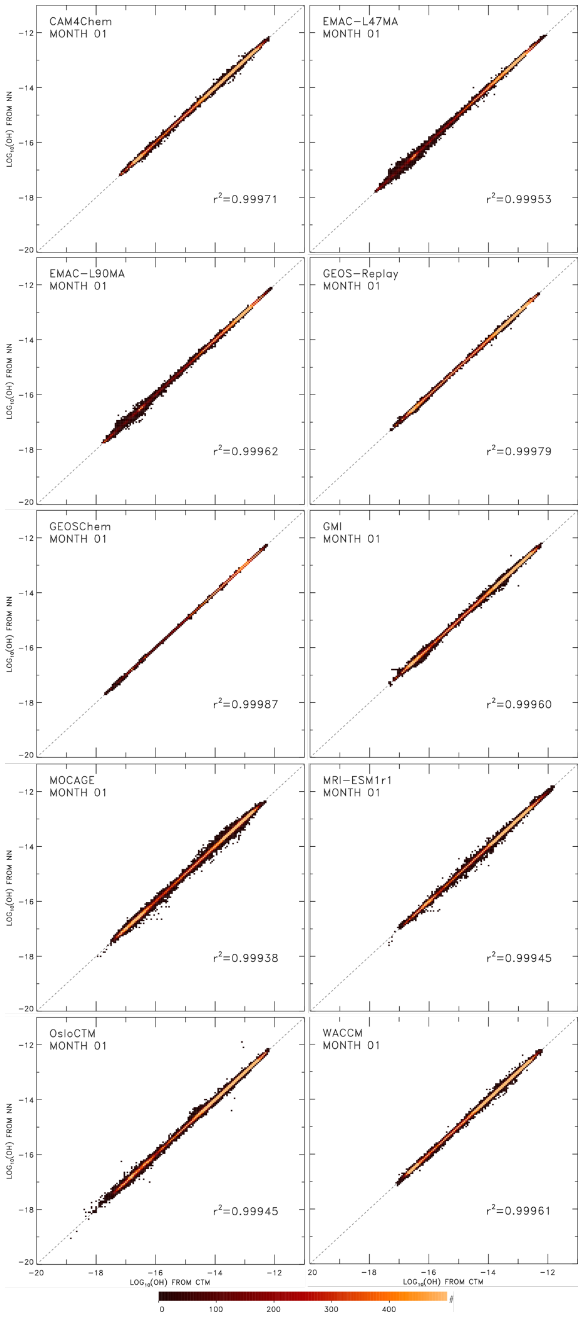


Figure S1. LOG$_{10}$ of OH mixing ratio calculated by NN versus that simulated by the native global model, for January. Colors indicate 2-D histogram to give indication of density of data points. All tropospheric model grid points are shown, including training, validation, and test data. Native model is indicated in the upper left corner of each panel, r^2 value is inscribed in the lower right, and 1:1 line is depicted as dashed grey line.

Formatted: Font: Bold

Formatted: Font: Bold

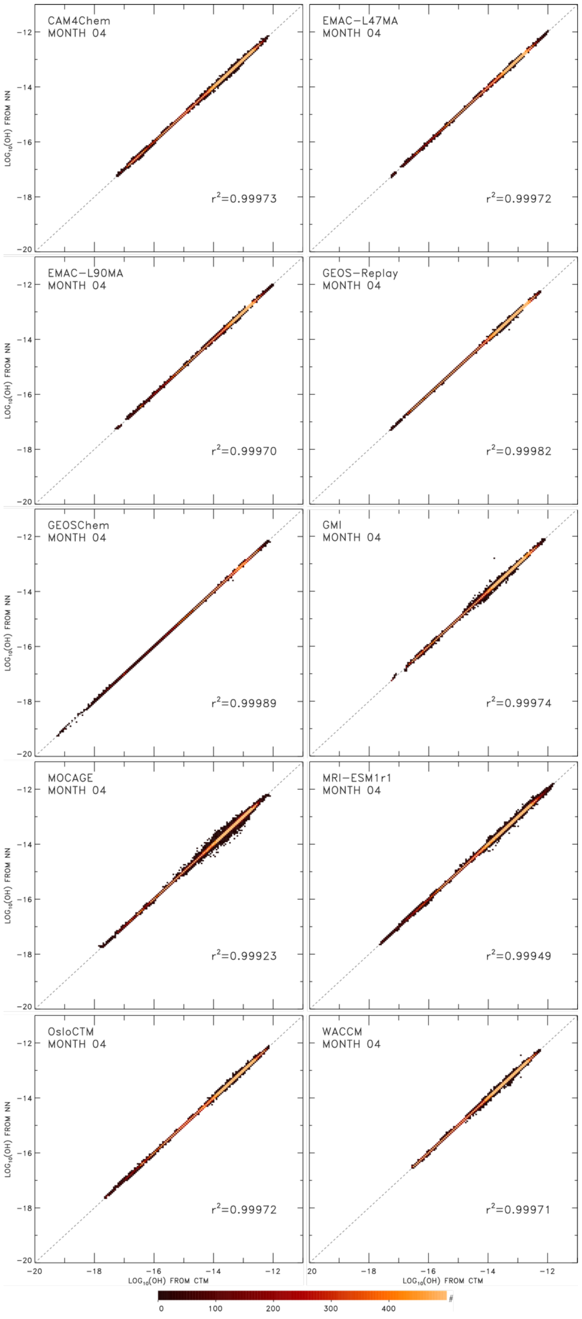


Figure S2. Same as Figure S1 except for month of April.

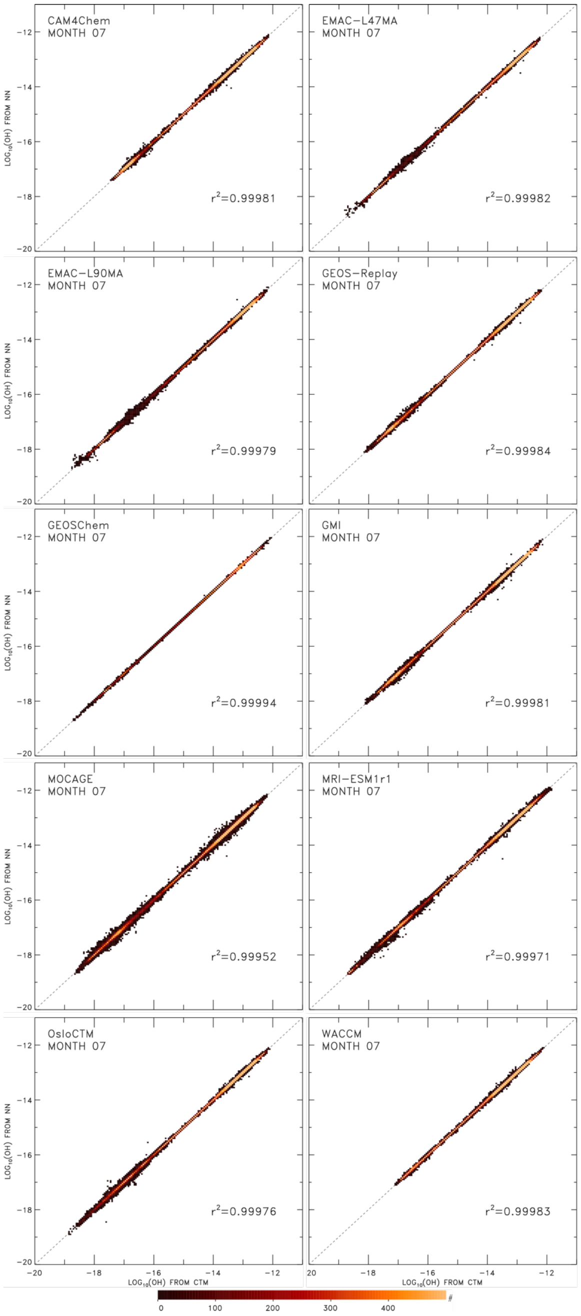


Figure S3. Same as Figure S1 except for month of July.

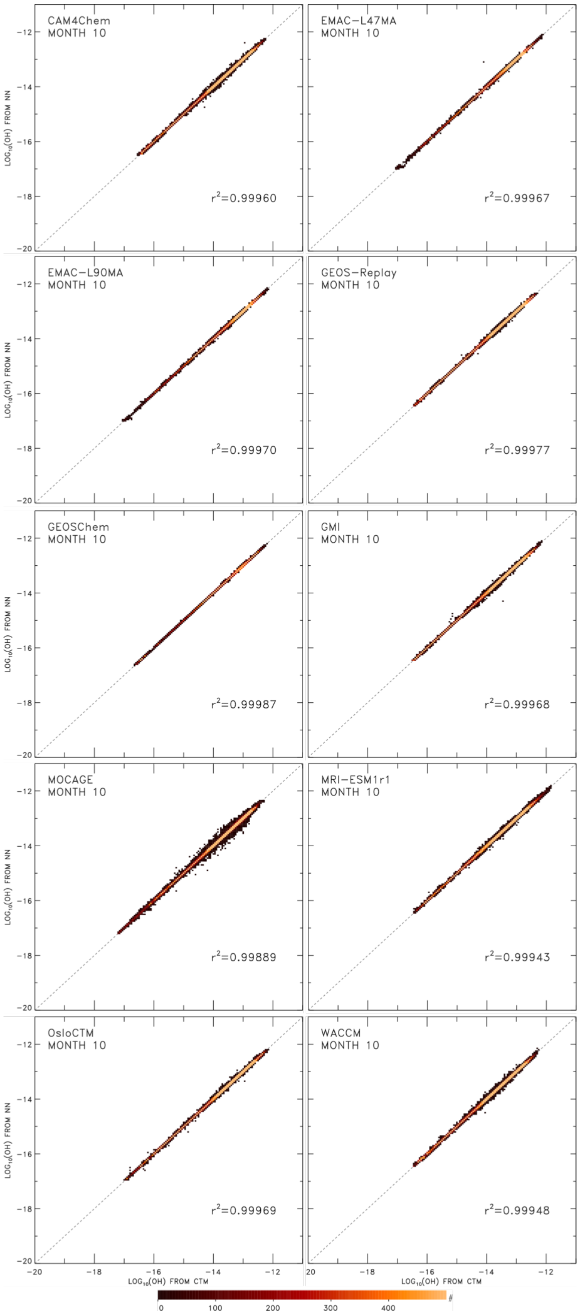


Figure S4. Same as Figure S1 except for month of October.

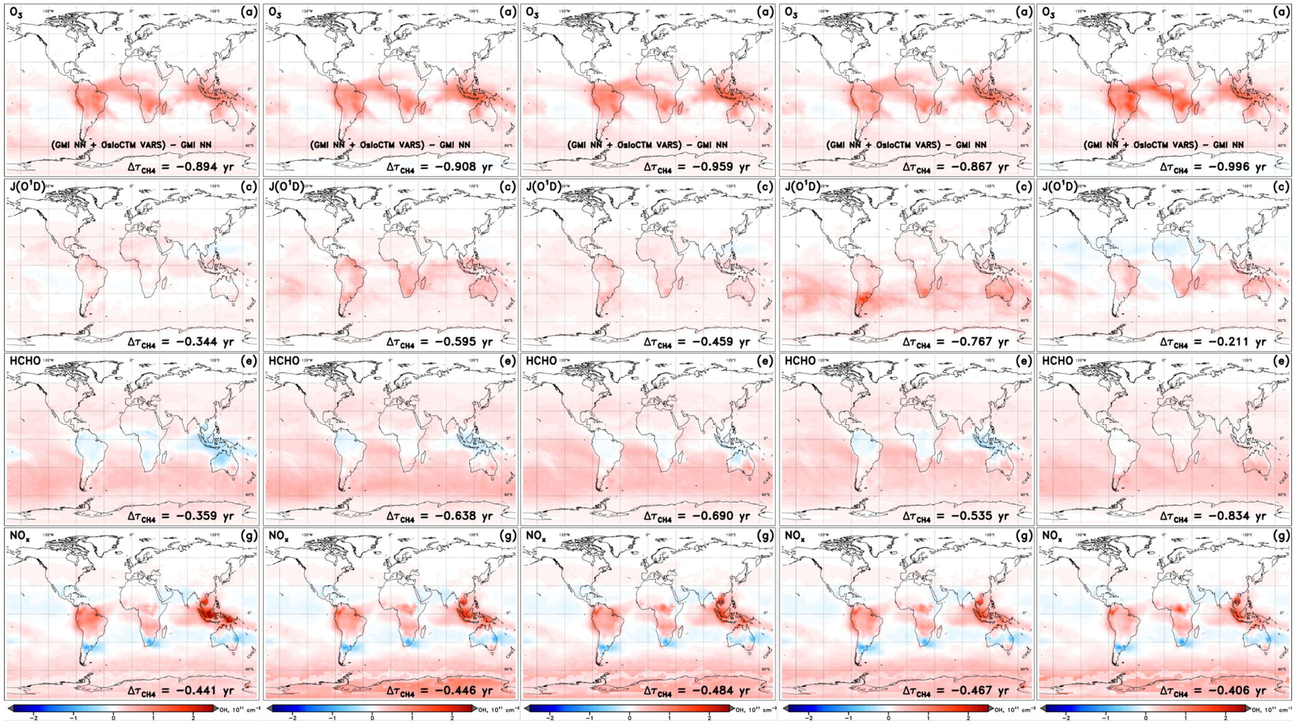


Figure S5. The equivalent of the left panels from Fig. 4 in the main text, evaluated for the other trainings of the January GMI NN. The second training (in the second column) produced the NN chosen for subsequent analysis. Training iteration is indicated by the column; the species swapped from the OsloCTM model into the GMI NN is indicated by the row (O_3, J(O^1D), HCHO, and NO_x, from top to bottom).

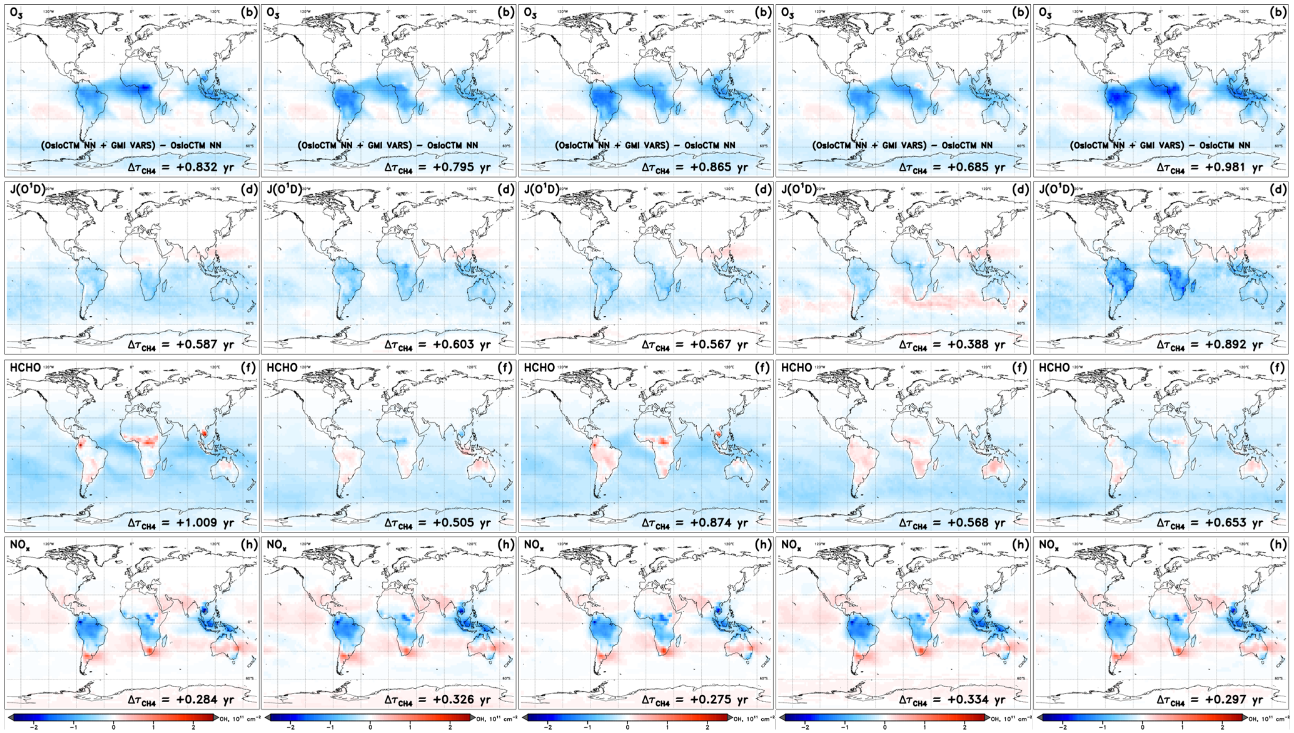
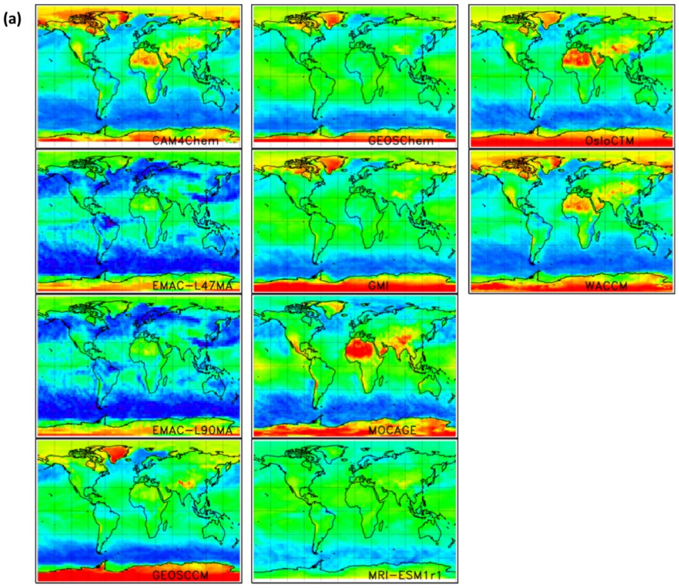


Figure S6. Same as Figure S5 except for the five trainings of the OsloCTM NN for January. The training instance for the January OsloCTM NN chosen for use in the broader analysis of this work was likewise the second, located in the second column.

Specified Dynamics, REF-C1SD



Free-running, REF-C1

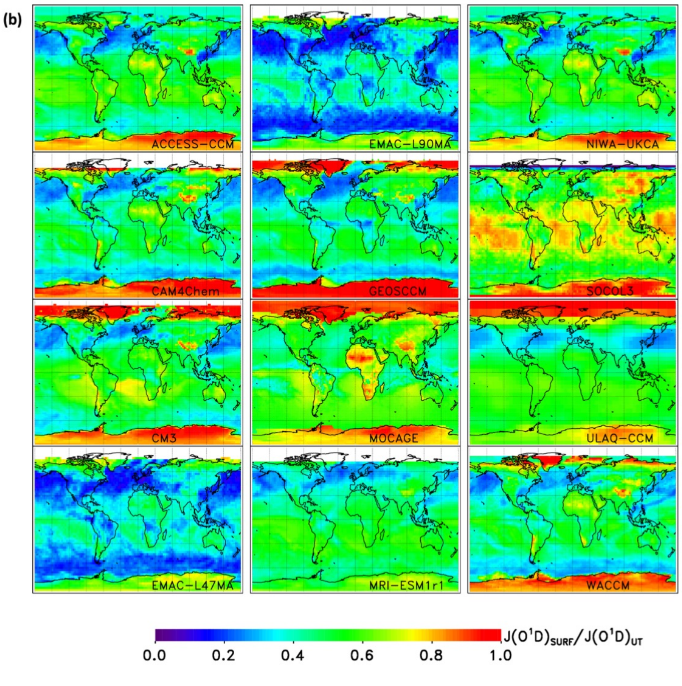


Figure S7, The ratio of JO^{1}D at the surface to JO^{1}D at the last pressure level within the troposphere before crossing the tropopause for the month in which each simulation set exhibited the largest τ_{CH4} differences attributed to JO^{1}D. (a) shows results from the REF-C1SD simulations for the month of April; (b) shows the REF-C1 simulations for the month of January. Suppression of this ratio below 1.0 is expected to result from cloud cover or other forms of absorption or scattering (by tropospheric O$_3$, aerosols, etc.).

Deleted: 1

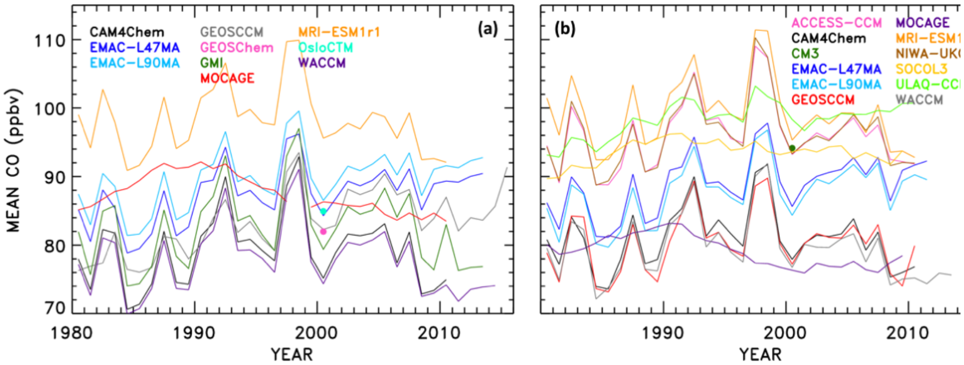


Figure S8. Time series of annually averaged CO mixing ratios at pressures greater than 700 hPa and latitudes between 30°S and 30°N from (a) the specified dynamics REF-C1SD simulations and (b) the free-running REF-C1 simulations.

Deleted: ¶

Page Break

¶

Formatted: Font: Not Bold

Deleted: 2

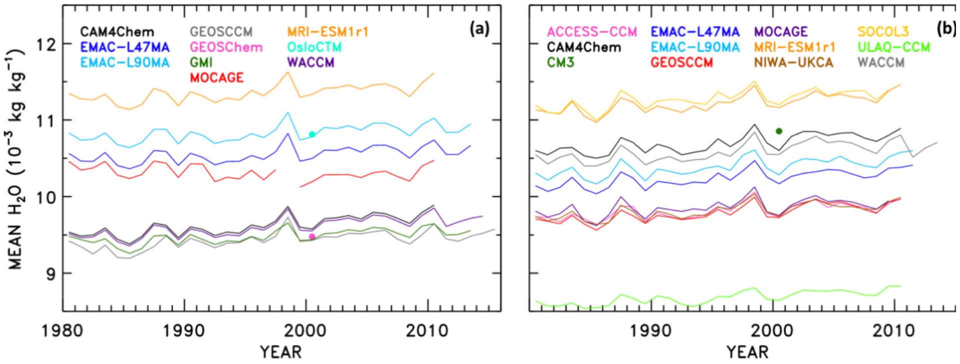


Figure S9. Time series of annually averaged H_2O mixing ratios at pressures greater than 700 hPa and latitudes between 30°S and 30°N from (a) the specified dynamics REF-C1SD simulations and (b) the free-running REF-C1 simulations.

Deleted: 3

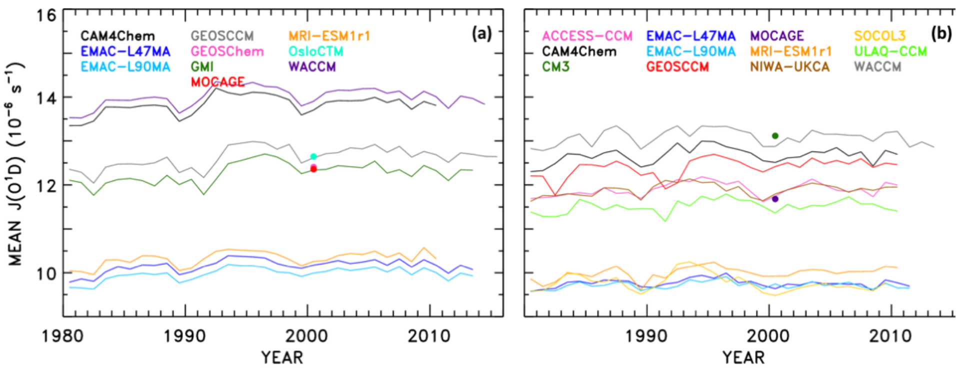


Figure S10. **Time series of annually averaged J(O^{1}D) frequencies at pressures greater than 700 hPa and latitudes between 30°S and 30°N from (a) the specified dynamics REF-C1SD simulations and (b) the free-running REF-C1 simulations.**

Deleted: 4

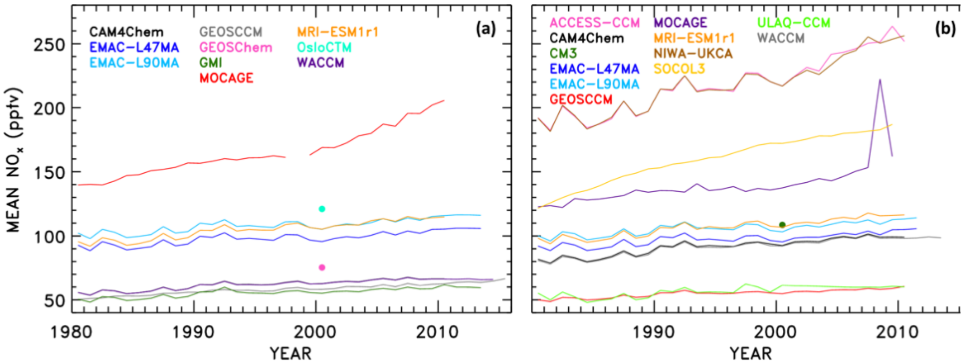


Figure S11. **Time series of annually averaged NO$_x$ mixing ratios at pressures greater than 700 hPa and latitudes between 30°S and 30°N from (a) the specified dynamics REF-C1SD simulations and (b) the free-running REF-C1 simulations.**

Deleted: 5

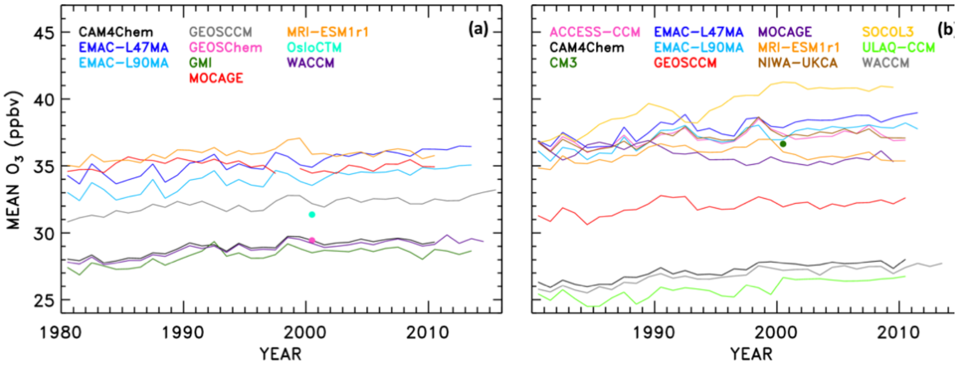


Figure S12. Time series of annually averaged O_3 mixing ratios at pressures greater than 700 hPa and latitudes between 30°S and 30°N from (a) the specified dynamics REF-C1SD simulations and (b) the free-running REF-C1 simulations.

Deleted: 6

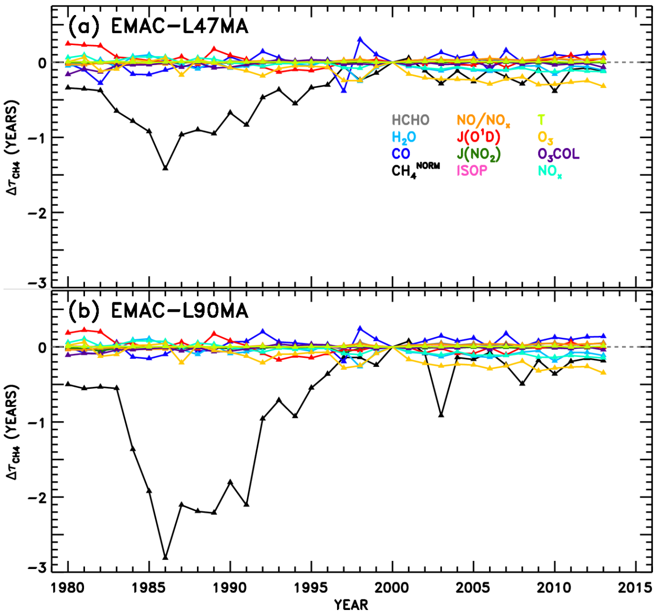


Figure S13. As in Figure 7 of the main text, panels (b) and (c), but with expanded y-axes to show the full range of the τ_{CH_4} responses to CH_4^{NORM} from the two EMAC model configurations.

Formatted: Centered

Formatted: Font: 9 pt

Formatted: Font: 9 pt

Formatted: Font: 9 pt

Formatted: Font: Bold

Formatted: Font: 9 pt

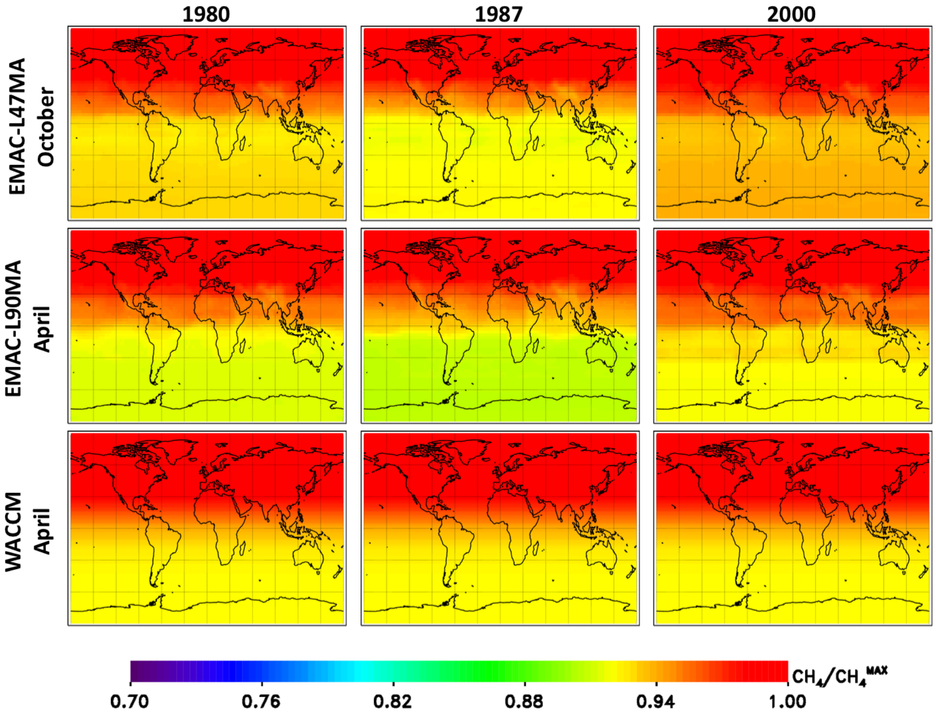


Figure S14. The ratio of local CH$_4$ mixing ratio to the maximum CH$_4$ mixing ratio found in the troposphere of a given model-simulated month, visualized for the pressure level nearest the surface, for the models, months, and years indicated. Results shown are for the REF-C1SD simulations. The normalized CH$_4$ quantity is used as an input to the neural networks to avoid issues introduced by non-overlapping fields of CH$_4$ absolute values between models and between years. This scaled CH$_4$ quantity is thus more accurately described as a measure of the CH$_4$ distribution within the troposphere. While the CH$_4$ distribution remains near-constant from year to year for a given month for most models (e.g., WACCM, bottom), the two configurations of the EMAC model show deviations from the trained (year 2000) distribution. Most notably between the mid-1980s and mid-1990s, CH$_4$ in the Southern Hemisphere decreases, relative to the higher CH$_4$ values in the Northern Hemisphere. It is these deviations in the EMAC CH$_4$ distributions that are likely driving the anomalous τ_{CH4} response in Fig. 7.

Deleted: 7

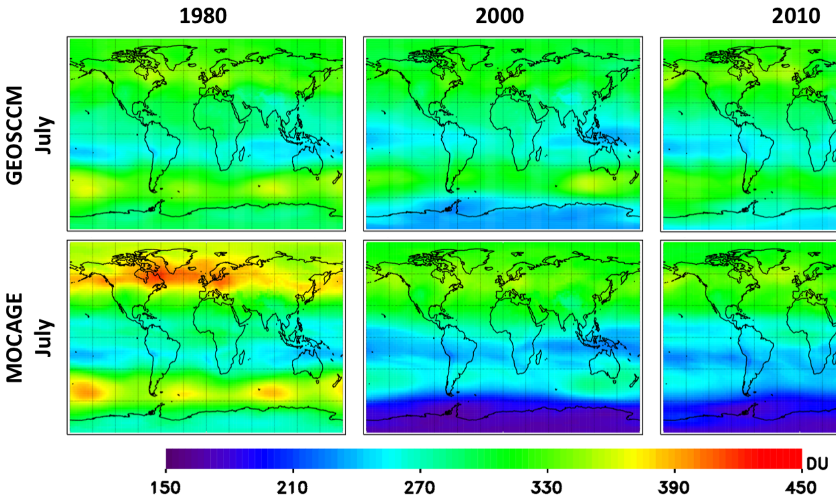


Figure S15. July total ozone columns from the GEOSCCM (top) and MOCAGE (bottom) REF-C1SD simulations for year 1980 (left), 2000 (center), and 2010 (right). While model differences between GEOSCCM and MOCAGE are apparent, it is the stark difference between year 1980 and 2000 in the MOCAGE model that is concerning and likely driving the anomalous τ_{CH4} response in the early- to mid-1980s, seen in Fig. 7. Other models, such as GEOSCCM, do not show such drastic differences between year 1980 and 2000 ozone column amounts.

Deleted: 8

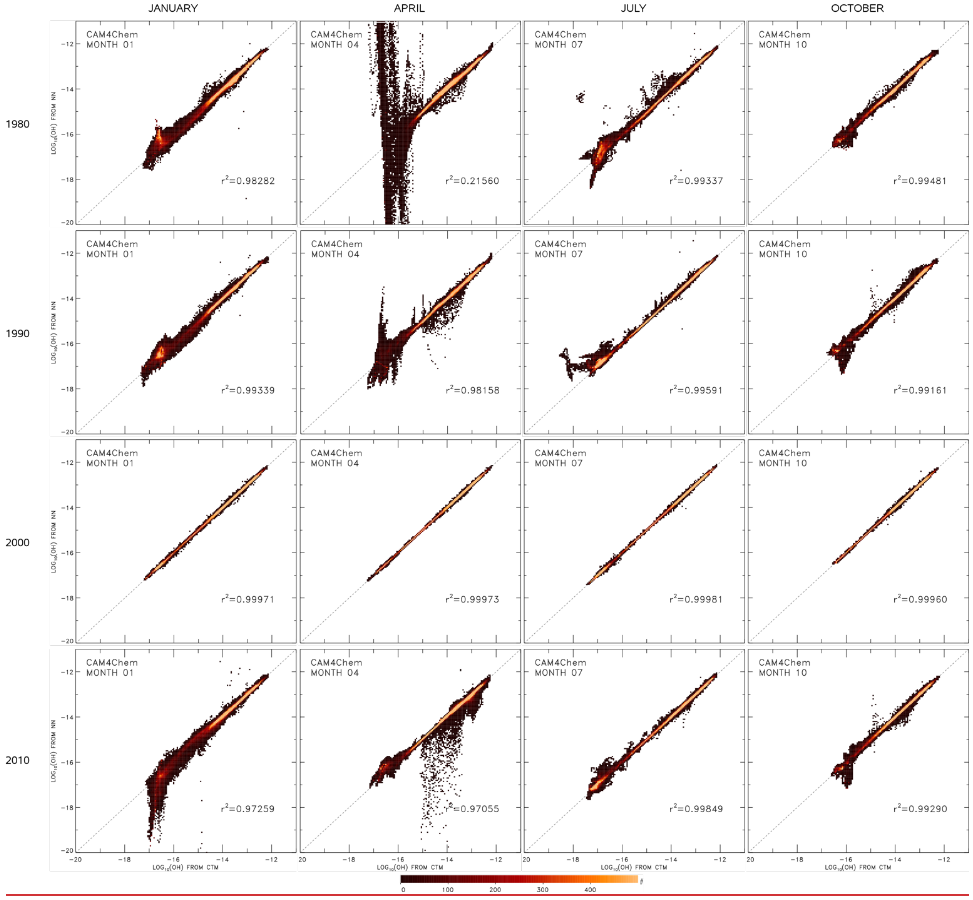


Figure S16. NN performance plots, as in Figure S1, but for only the REF-C1SD simulation of the CAM4-Chem model, evaluated for years 1980, 1990, 2000, and 2010 (from top to bottom). For each month (represented by the column), the same NN, trained on year 2000 model output, is subsequently run with inputs taken from the alternative years.

Formatted: Font: Bold

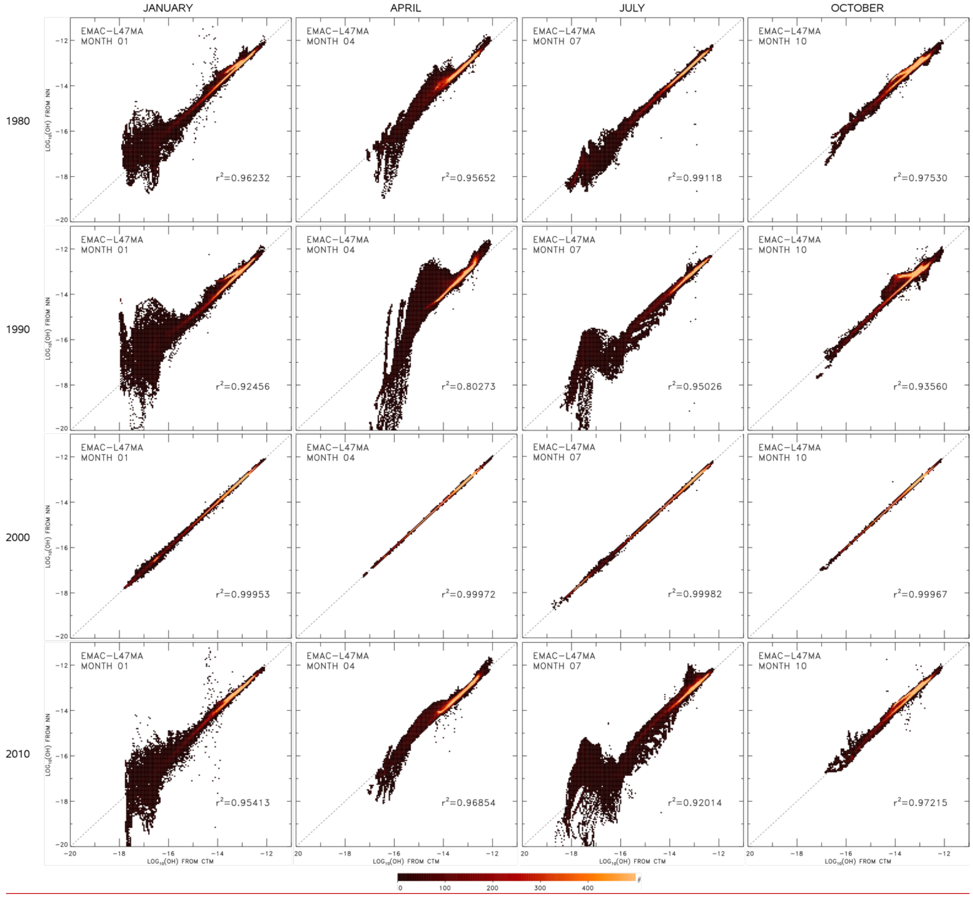


Figure S17. As in Figure S16, but for the REF-C1SD simulation of the EMAC-L47MA model.

Formatted: Font: Bold

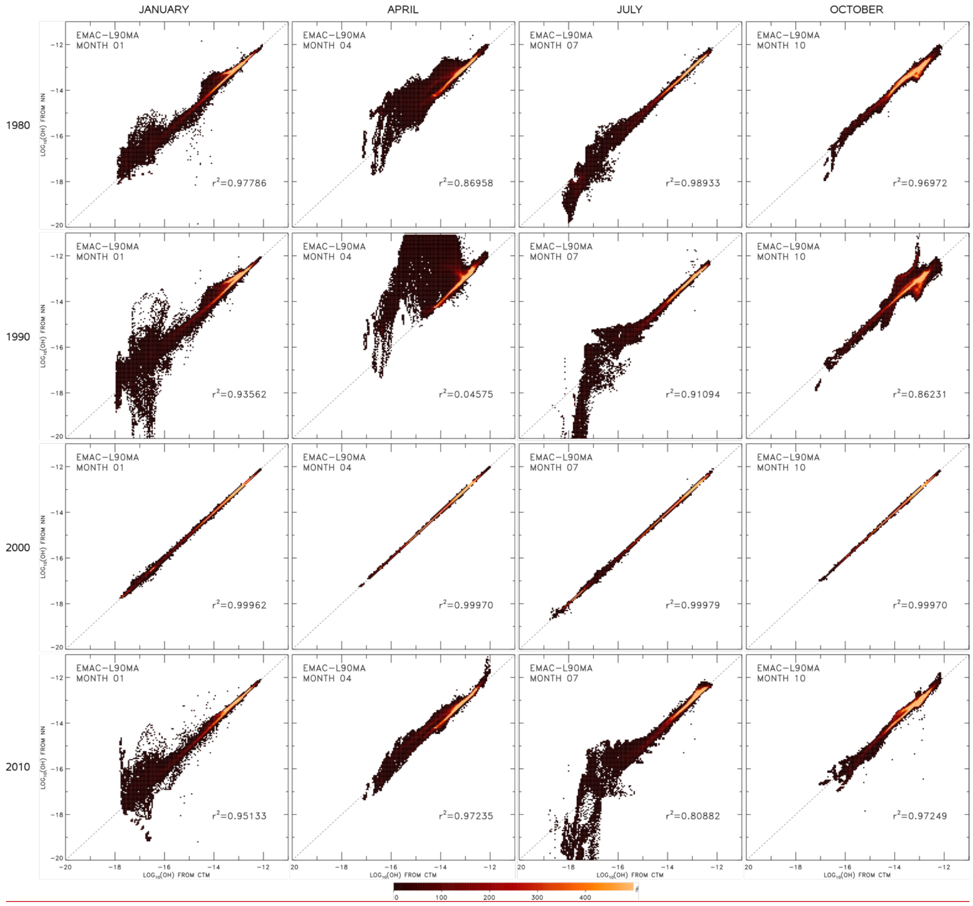


Figure S18. As in Figure S16, but for the REF-C1SD simulation of the EMAC-L90MA model.

Formatted: Font: Bold

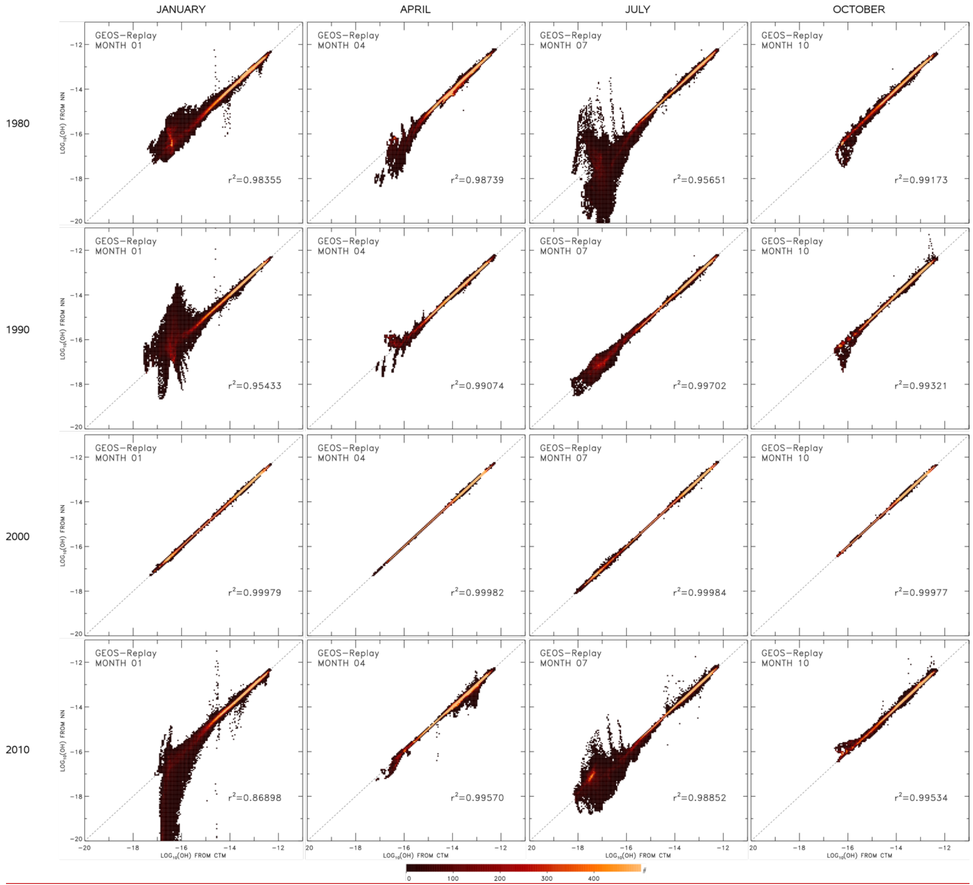


Figure S19. As in Figure S16, but for the REF-C1SD simulation of the GEOS-Replay model.

Formatted: Font: Bold

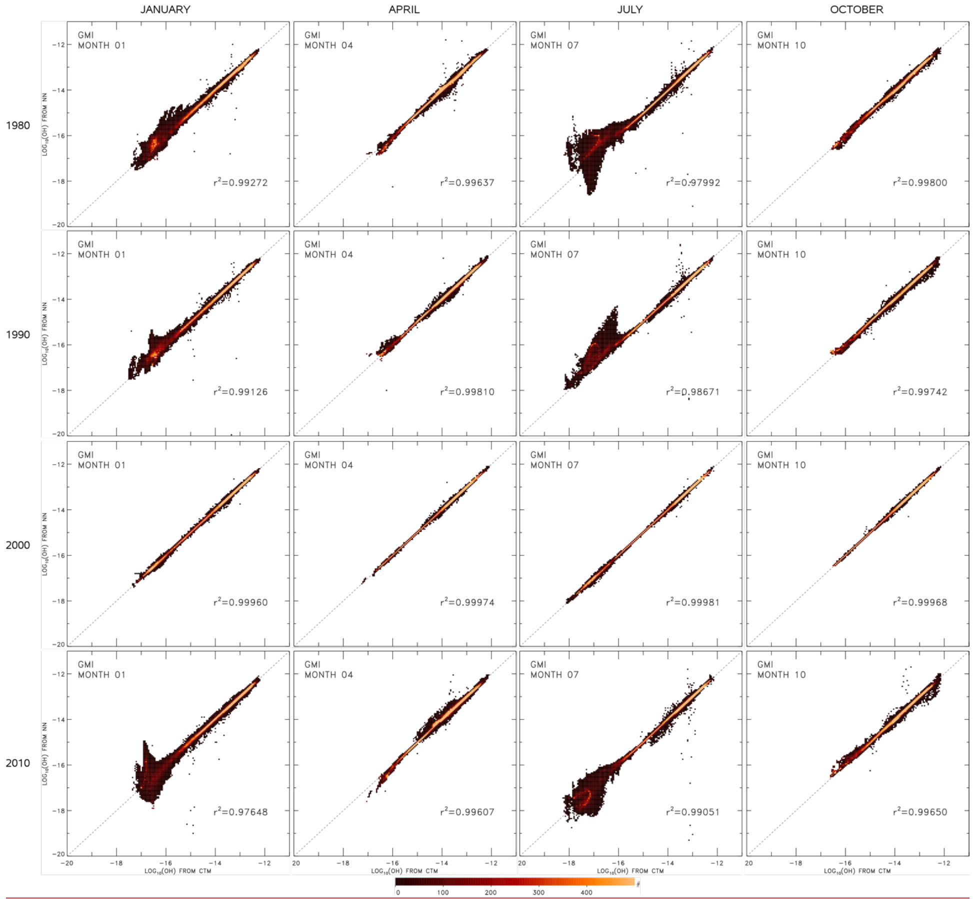


Figure S20. As in Figure S16, but for the REF-C1SD simulation of the GMI model.

Formatted: Font: Bold

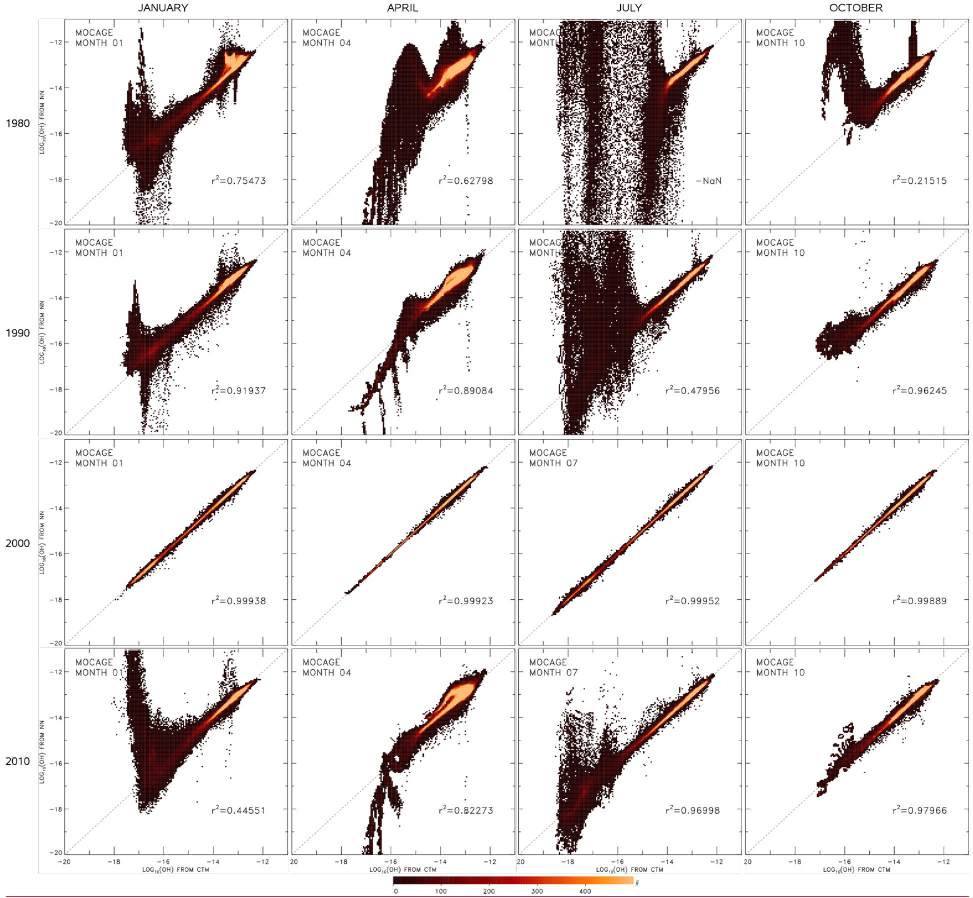


Figure S21. As in Figure S16, but for the REF-C1SD simulation of the MOCAGE model.

Formatted: Font: Bold

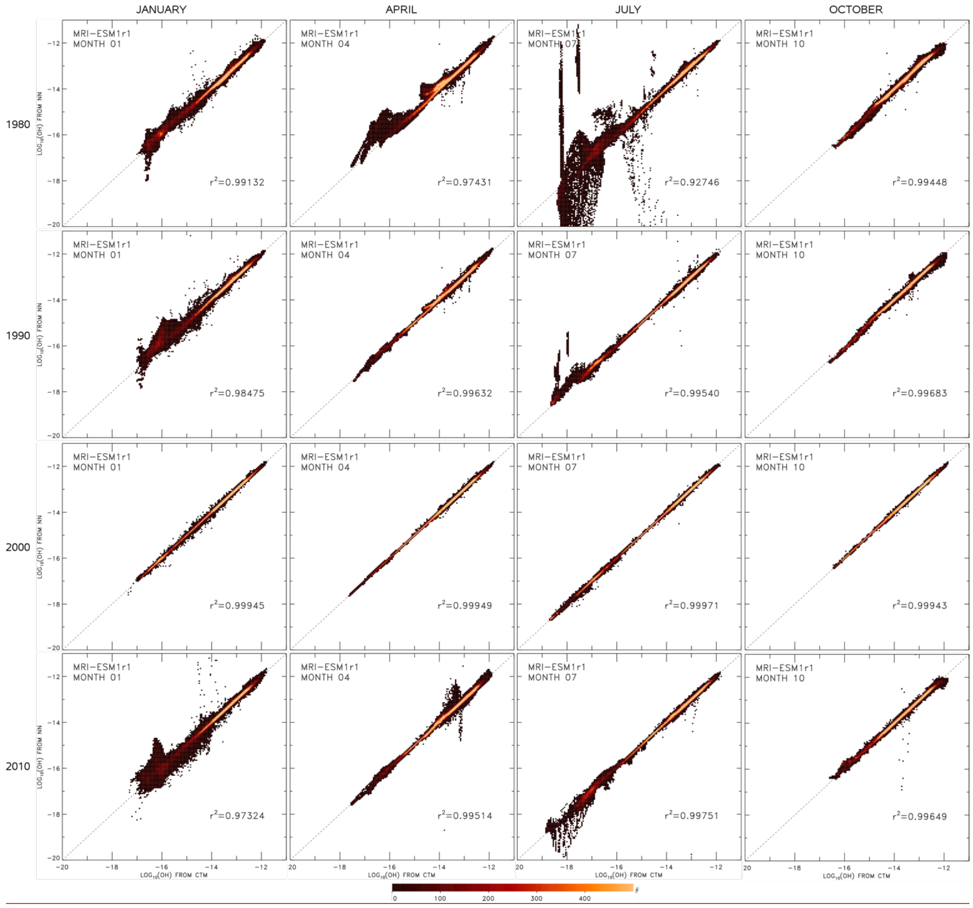


Figure S22. As in Figure S16, but for the REF-C1SD simulation of the MRI-ESM1r1 model.

Formatted: Font: Bold

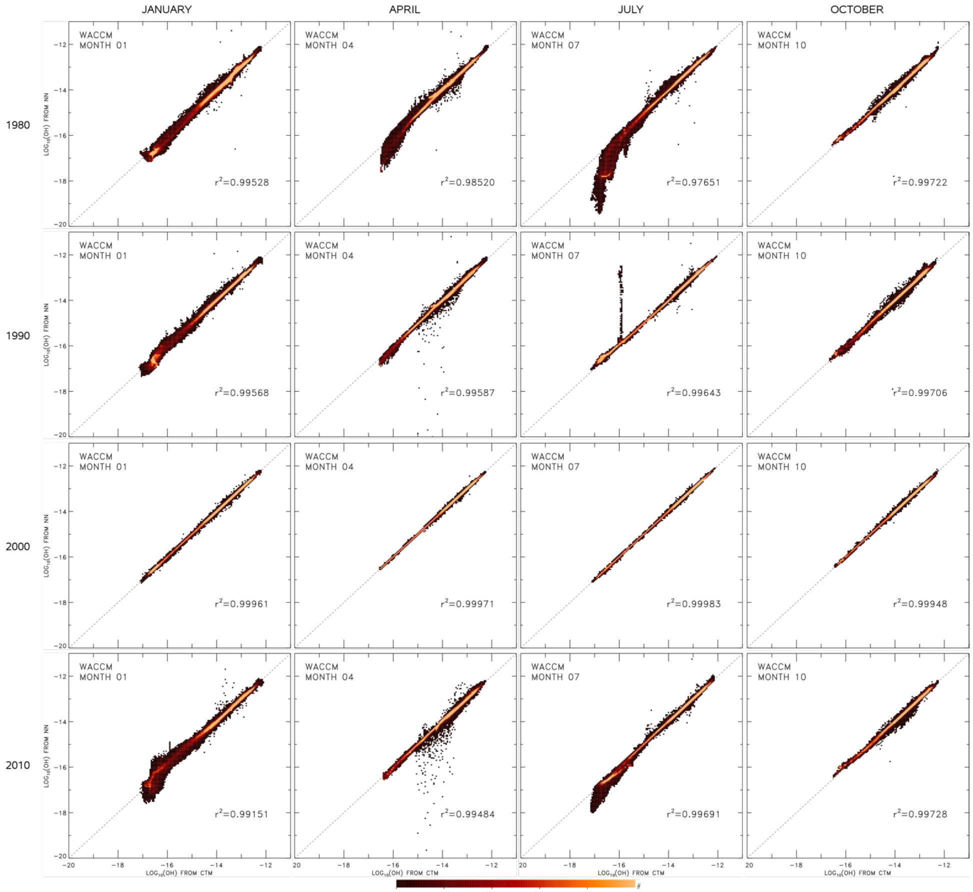


Figure S23. As in Figure S16, but for the REF-C1SD simulation of the WACCM model.

Formatted: Font: Bold

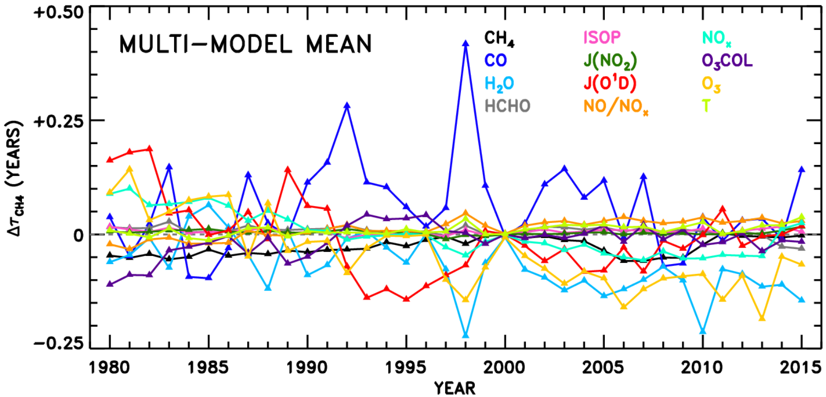


Figure S24. As in Figure 8 of the main text, but without application of the NN quality check described in Section 4.3.1. In other words, all NN results, from all models and years shown in Fig. 7 of the main text, for the attribution of τ_{CH_4} changes are included in the multi-model mean, except for the three cases that stood out by eye as spurious: EMAC-L47MA CH_4, EMAC-L90MA CH_4, and MOCAGE O_3 COL.

Formatted: Centered

Formatted: Font: 9 pt

Formatted: Font: 9 pt

Formatted: Font: 9 pt

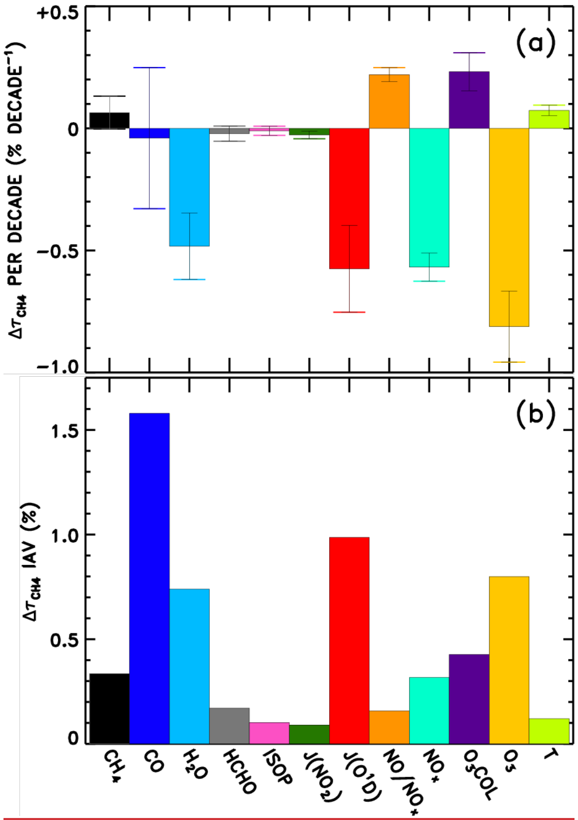


Figure S25. As in Figure 9 of the main text, but without application of the NN quality check described in Section 4.3.1.

Formatted: Centered

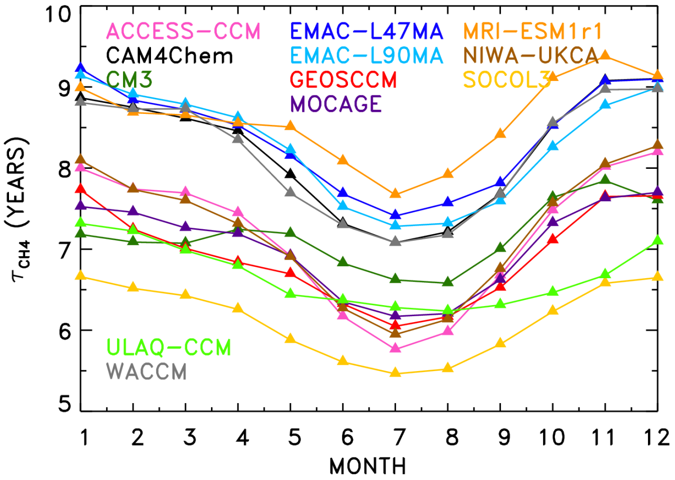


Figure S26. Seasonal variation in CH_4 lifetime for year 2000 for the CCMI free-running (REF-C1) model simulations.

Deleted: 9

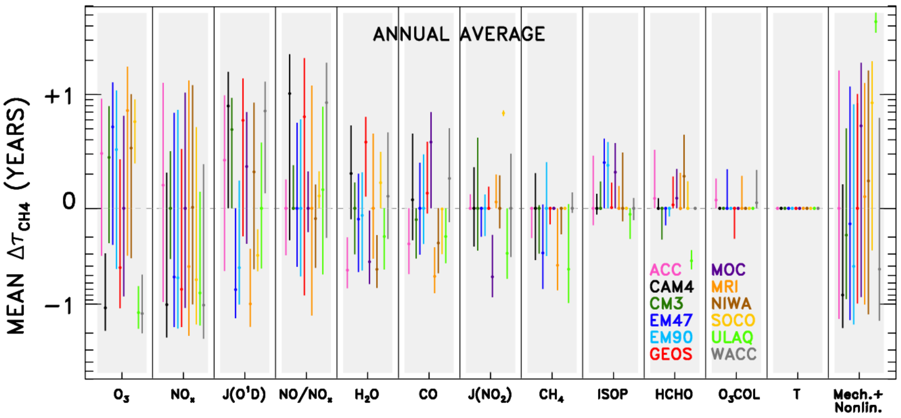
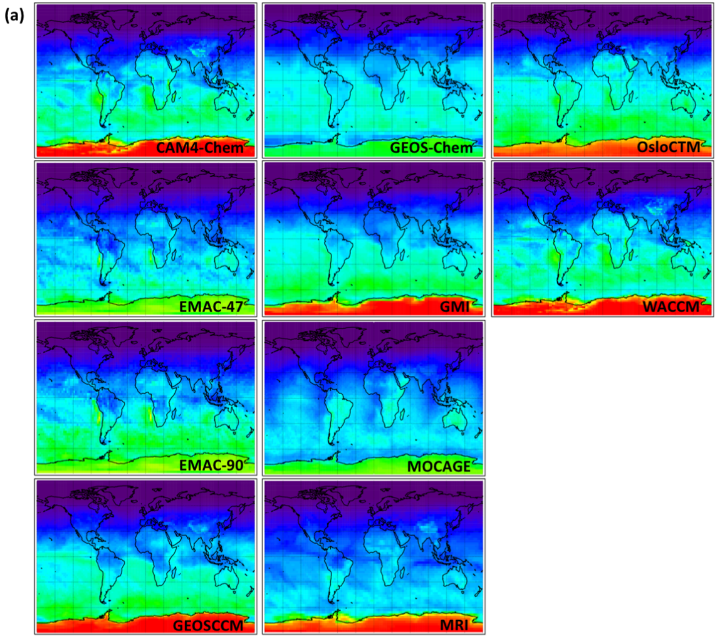
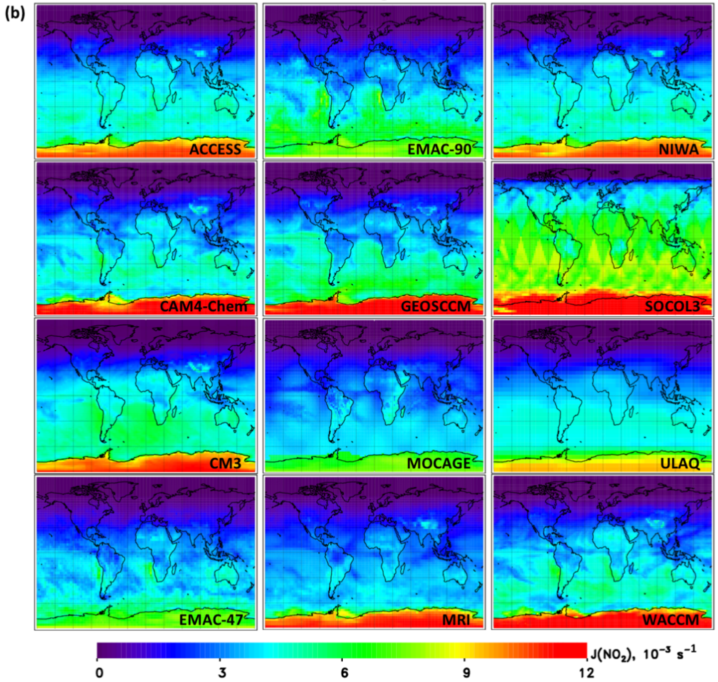


Figure S27. Averaged changes in CH_4 lifetime for the free-running (REF-C1) CCMI simulations. Values of $\Delta\tau_{CH4}$ are accrued for a specified model (color), across all swaps of the indicated variable (x-axis) from all other models. Results are shown annually averaged for year 2000 of the specified dynamics REF-C1SD CCMI and chemical transport model simulations. Circle indicates the mean change in CH_4 lifetime; bars represent the 1σ standard deviation from all model pairings. Variables along the x-axis are ranked by averaged magnitude of the $\Delta\tau_{CH4}$ values (i.e., inputs located farther left are responsible for larger differences in CH_4 lifetime), except for the "Mech.+Nonlin." term, which is shown last to indicate its role as a remainder term.

Deleted: 10



Specified Dynamics, REF-C1SD



Free-running, REF-C1

Figure S28. JNO$_2$ values directly from each model at the pressure level closest to 850 hPa for January, 2000 of (a) the REF-C1SD simulations and (b) the REF-C1 simulations.

Deleted: 11